

Search completed: May 6, 2004, 09:30:46  
 Job time : 52.724 secs

XX WO200252015-A2.  
 XX PD 04-JUL-2002.  
 XX PF 20-DEC-2001; 2001WO-CA001843.  
 XX PR 22-DEC-2000; 2000US-0257857P.  
 XX DR N-PSDB; ABX88587.  
 PA (BOEHRINGER INGELHEIM CANADA LTD.  
 XX PT New self-replicating RNA molecules from Hepatitis C virus (HCV), which  
 PT possess enhanced transduction or replication efficiency, useful for  
 PT evaluating potential inhibitors of HCV replication.  
 XX Disclosure; Page 120-129; 140PP; English.  
 XX  
 CC The invention describes a self-replicating hepatitis C virus (HCV)  
 CC polynucleotide molecule comprising a 5'-non translated region (5'NTR),  
 CC where guanine at position 1 is substituted for adenine, a HCV polyprotein  
 CC region coding for a HCV polyprotein, and a 3'-NTR region. The self-  
 CC replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating  
 CC potential inhibitors of HCV replication. The HCV RNA molecule is also  
 CC useful for efficiently establishing cell culture replication. The self-  
 CC replicating polynucleotide molecule contains a 5'-NTR, where G at  
 CC position 1 is substituted for A, and therefore provides an alternative to  
 CC existing systems comprising a self-replicating HCV RNA molecule that, in  
 CC conjunction with mutations in the HCV non-structural region, such as the  
 CC G(2042)C/R mutations, transduces and/or replicates with greater  
 CC efficiency. This amino acid sequence is encoded by the hepatitis C virus  
 CC replicon APGK2 and contains the viral protease NS3/3, protease complex  
 CC NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B  
 XX Sequence 2201 AA;  
 SQ  
 Query Match 95.0%; Score 1887; DB 5; Length 2201;  
 Best Local Similarity 93.4%; Pred. No. 9.9e-14; Length 2201;  
 Matches 354; Conservative 13; Mismatches 12; Indels 0; Gaps 0;  
 QY  
 2 LMLTSPYYKVLALIWWLQYLITRVEAHQLQWMPPLNVRGERDAATLILTCVHPPELFID 61  
 Db 19 LLLTSLSPHYKLFLARLUWWLQYFTRAEAHQLQWMPPLNVRGERDAVILTCVHPPELFIT 78  
 QY 62 ITKLLAIGPMLIQAGITKVPFVRAQGLTACMVLRAKGAGHYVOMAFKLAALTGT 121  
 Db 79 ITKLLAIGPMLIQAGITKVPFVRAQGLTACMVLRAKGAGHYVOMAFKLAALTGT 138  
 QY 122 YVPLHLPQDWAHGLRDIAVAEPVITSMEKILNGAGTAACGDIISGLPVSRGG 181  
 Db 139 YVVDILTPLRDWAGLRLSLAVAVEPWVHSMDMERTKUTWGAATACGDIILGLPVSRGG 198  
 QY 182 REILLGPADNFEGQWRLIAPITASQQTQGLGCGISLTGRDKNOEGEGQVNSTATQ 241  
 Db 199 REIHGPASLLEGQWRLIAPITASQQTQGLGCGISLTGRDKNOEGEGQVNSTATQ 258  
 QY 242 SFLATCVCNGCWTVPFHGAGSKTLAQPKGPTQMYTNVDLQWQAPPGARSMTPCTCGS 301  
 Db 259 SPLATCVCNGCWTVPFHGAGSKTLAQPKGPTQMYTNVDLQWQAPPGARSMTPCTCGS 318  
 QY 302 SDLYNTRHADWVIPVRRGSRGSLSLSPREVSYIKGSSGQPLLCPSGHAVGIFRAAVCTR 361  
 Db 319 SDLYNTRHADWVIPVRRGSDRSRGSLSLSPREVSYIKGSSGQPLLCPSGHAVGIFRAAVCTR 378  
 QY 362 GVAKAVDFPESMETMR 380  
 Db 379 GVAKAVDFPESMETMR 397



|  |  |   |        |         |         |   |  |
|--|--|---|--------|---------|---------|---|--|
| Copyright (c) 1993 - 2004 Compugen Ltd.  | GenCore version 5.1.6  |   |        |         |         |   |  |
| OM protein - protein search, using sw model  |  |   |        |         |         |   |  |
| Run on:  | May 6, 2004, 09:22:36 ; Search time 12.3758 Seconds          |   |        |         |         |   |  |
| Title:   | US-10-650-585-12   |   |        |         |         |   |  |
| Perfect score:   | 1987   |   |        |         |         |   |  |
| Sequence:  | 1 ALLTISPVYKVLILLNLWLL.....RGVAKAVDIFIPVESMETMR 380          |   |        |         |         |   |  |
| Scoring table:   | BLOSUM62   |   |        |         |         |   |  |
| Searched:  | Gapop 10.0 , Capext 0.5                                      |   |        |         |         |   |  |
| Total number of hits satisfying chosen parameters:   | 283366   |   |        |         |         |   |  |
| Minimum DB seq length: 0   |  |   |        |         |         |   |  |
| Maximum DB seq length: 2000000000  |  |   |        |         |         |   |  |
| Post-processing:   | Minimum Match 0%<br>Maximum Match 100%                       |   |        |         |         |   |  |
| Listing first 45 summaries   |  |   |        |         |         |   |  |
| Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution. |  |   |        |         |         |   |  |
| database :   | PIR 78:<br>1: pirl:<br>2: pir2:<br>3: pir3:<br>4: pir4:<br>* |   |        |         |         |   |  |
| <b>SUMMARIES</b>   |  |   |        |         |         |   |  |
| Result No.   | Score  | Query   | Length | DB      | ID      | Description   |  |
| 1  | 1902   | 95.7  | 3010   | 1       | GNWVCJ  | genome polyprotein - hepatitis C virus (strain J)   |  |
| 2  | 1890   | 95.1  | 3010   | 1       | A45573  | genome polyprotein  |  |
| 3  | 1870   | 94.1  | 3010   | 1       | GNWVWT  | genome polyprotein  |  |
| 4  | 1845   | 92.9  | 3010   | 1       | S18030  | genome polyprotein  |  |
| 5  | 1823   | 91.7  | 3010   | 1       | GNWVTC  | genome polyprotein  |  |
| 6  | 1715   | 86.3  | 3011   | 1       | GNWVC3  | genome polyprotein  |  |
| 7  | 1709   | 86.0  | 3011   | 1       | S40770  | genome polyprotein  |  |
| 8  | 1703   | 85.7  | 3011   | 1       | GNWCH   | genome polyprotein  |  |
| 9  | 1482   | 74.6  | 3014   | 1       | JCS5620 | genomic RNA   |  |
| 10   | 1391   | 70.0  | 3033   | 1       | JCL1303 | Residues: 1-3010 <KAT>  |  |
| 11   | 1391   | 70.0  | 3033   | 1       | GNWVJB  | A;Cross-references: GB:0920208; NID:9221610; PIDN:BA14233.1; PID:9221611                      |  |
| 12   | 408.5  | 20.6  | 3005   | 2       | T08841  | R;Kato, N.; Ohkoshi, S.; Shimotohno, K.   |  |
| 13   | 342.5  | 17.2  | 2970   | 2       | T08839  | Proc. Jpn. Acad. 65B, 219-223, 1989   |  |
| 14   | 112  | 5.6   | 592    | 2       | H71425  | A;Title: Japanese isolates of the non-A, non-B hepatitis viral genome show sequence variation |  |
| 15   | 102.5  | 5.2   | 660    | 1       | VHWWHH  | A;Reference number: A39253  |  |
| 16   | 101  | 5.1   | 564    | 2       | S36637  | A;Accession: PS0086   |  |
| 17   | 101  | 5.1   | 600    | 2       | S46642  | A;Molecule type: genomic RNA  |  |
| 18   | 100.5  | 5.1   | 353    | 2       | G87392  | A;Residues: 1-2707 <KAT>  |  |
| 19   | 99   | 5.0   | 399    | 2       | AH3038  | A;Experimental source: Japanese isolate   |  |
| 20   | 99   | 5.0   | 399    | 2       | C98247  | C;Comment: The cleavage sites of this polyprotein have not been determined.                   |  |
| 21   | 97.5   | 4.9   | 1085   | 2       | T03531  | C;Superfamily: hepatitis C virus genome polyprotein   |  |
| 22   | 95.5   | 4.8   | 470    | 2       | JCC4098 | C;Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serine     |  |
| 23   | 94.5   | 4.8   | 2796   | 2       | JCC4743 | F-2-115/Product: capsid protein C #status predicted <>PC>                                     |  |
| 24   | 93.5   | 4.7   | 446    | 2       | AP1509  | F-116-151/Product: envelope protein M #status predicted <ERM>                                 |  |
| 25   | 93.5   | 4.7   | 1380   | 2       | T18309  | F-192-389/Product: major envelope protein E #status predicted <ME>                            |  |
| 26   | 93   | 4.7   | 7463   | 2       | T36248  | F-390-729/Product: nonstructural protein NS1 #status predicted <NS1>                          |  |
| 27   | 92.5   | 4.7   | 706    | 2       | S33761  | F-730-1006/Product: nonstructural protein NS2 #status predicted <NS2>                         |  |
| 28   | 92.5   | 4.7   | 716    | 2       | G83612  | F-1007-1615/Product: hepacivirin #status predicted <NS3>                                      |  |
| 29   | 92.5   | 4.7   | 3069   | 2       | H70656  | F-1230-1237/Region: nucleotide-binding motif A (P-loop)                                       |  |
| <b>ALIGNMENTS</b>  |  |   |        |         |         |   |  |
| Query Match  | 95.7%  | Score   | 1902   | DB      | 1       | Length  | 3010;  |
| Best Local Similarity  | 94.2%  | Pred. No.   | 6      | 9e-150; |         | Matches   | 357; Conservative 12; Mismatches 10; Indels 0; Gaps 0; |
| Oy   | 2  | LILTISPVYKVLILLNLWLL.....RGVAKAVDIFIPVESMETMR       | 61     |         |         |   |  |
| Db   | 828  | LILTISPVYKVLILLNLWLL.....RGVAKAVDIFIPVESMETMR       | 887    |         |         |   |  |
| Oy   | 62   | ITKLLAATGPNLQLQAGIKTPYFQAQLIRACMVRKAGGHVYQAMKLAALGT | 121    |         |         |   |  |
| Db   | 888  | ITKLLAATGPNLQLQAGIKTPYFQAQLIRACMVRKAGGHVYQAMKLAALGT | 947    |         |         |   |  |

|  |  |  |      |          |      |  |      |
|--|--|--|------|----------|------|--|------|
| Qy   | 122  | YVVDHILPLODWAHAGLDRIDAVAVEPVIFSVDMEVKITTGADTAACGPIISGLPVSAARRG | 181  | Db       | 1008 | RRLIGPADSEQGMLRILITAYAQTCGLGIVTSIGRKQVQEVOVSTAQ                | 1067 |
| Db   | 948  | YVDLHILPLODWAHAGLDRIDAVAVEPVIFSVDMEVKITTGADTAACGPIISGLPVSAARRG | 1007 | Qy       | 242  | SFLATCNGVWTVFAGSKTLAPKPKPTOMYINVDLVGQAPPGARSMTPCTGS            | 301  |
| Qy   | 182  | REILGPGDFEGQWRLLAPITAYQSQRGLGICITSLGRDKNQVEGEVQVSTATQ          | 241  | Db       | 1068 | SFLATCNGVWTVFAGSKTLAPKPKPTOMYINVDLVGQAPPGARSMTPCTGS            | 1127 |
| Db   | 1008   | KELINGPDADEQWRLLAPITAYQSQRGLGICITSLGRDKNQVEGEVQVSTATQ          | 1067 | Qy       | 302  | SDLYLVRHADVIPRRRGSRGSLSPRIVSYLKGSGGPLCPSGHAGVIFRAAVCTR         | 361  |
| Qy   | 242  | SFLATCNGVWTVFAGSKTLAPKPKPTOMYINVDLVGQAPPGARSMTPCTGS            | 301  | Db       | 1128 | SDLYLVRHADVIPRRRGSRGSLSPRIVSYLKGSGGPLCPSGHAGVIFRAAVCTR         | 1187 |
| Db   | 1068   | SFLATCNGVWTVFAGSKTLAPKPKPTOMYINVDLVGQAPPGARSMTPCTGS            | 1127 | Qy       | 362  | GVAKAVDFIPESMETMR  | 380  |
| Qy   | 302  | SDLYLVRHADVIPRRRGSRGSLSPRIVSYLKGSGGPLCPSGHAGVIFRAAVCTR         | 361  | Db       | 1188 | GVAKAVDFIPESMETMR  | 1206 |
| Db   | 1128   | SDLYLVRHADVIPRRRGSRGSLSPRIVSYLKGSGGPLCPSGHAGVIFRAAVCTR         | 1187 | Qy       | 362  | GVAKAVDFIPESMETMR  | 380  |
| Qy   | 362  | GVAKAVDFIPESMETMR  | 380  | Db       | 1188 | GVAKAVDFIPESMETMR  | 1206 |
| Db   | 1188   | GVAKAVDFIPESMETMR  | 1206 | RESULT 2 |      |  |      |
| A;Residues:  | 1-3010 <TAN>   |  |      |          |      |  |      |
| A;Experimental source:   | RCV-JT   |  |      |          |      |  |      |
| A;Note:  | sequence extracted from NCBI backbone (NCBIn:116206, NCBI:P:106207)                                  |  |      |          |      |  |      |
| C;Species:   | hepatitis C virus  |  |      |          |      |  |      |
| C;Date:  | 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-Jan-2001                                  |  |      |          |      |  |      |
| C;Accession:   | A45573   |  |      |          |      |  |      |
| R;tanaka, T.; kato, N.; Nakagawa, M.; ootsuyama, Y.; cho, M.J.; Nakazawa, T.; hijikata, Virus Res. | 23, 39-53, 1992  |  |      |          |      |  |      |
| A;Title:   | Molecular cloning of hepatitis C virus genome from a single Japanese carrier.                        |  |      |          |      |  |      |
| A;Reference number:  | A45573; MUID:92295714; PMID:1318627  |  |      |          |      |  |      |
| A;Accession:   | A45573   |  |      |          |      |  |      |
| A;Molecule type:   | DNA  |  |      |          |      |  |      |
| A;Status:  | preliminary  |  |      |          |      |  |      |
| A;Cross-references:  | GB:D1168; GB:D01171; PID:BA01943_1; PID:g221513  |  |      |          |      |  |      |
| A;Note:  | host Homo sapiens (man)  |  |      |          |      |  |      |
| P;Chen, P.J.; Lin, M.H.; Tai, K.F.; Liu, P.C.; Lin, C.J.; Chen, D.S. Virology                      | 188, 102-113, 1992   |  |      |          |      |  |      |
| A;Title:   | The Taiwanese hepatitis C virus genome: sequence determination and mapping the : A;Reference number: | A40244; MUID:92230206; PMID:1314449                            |      |          |      |  |      |
| A;Accession:   | A40244   |  |      |          |      |  |      |
| A;Molecule type:   | genomic RNA  |  |      |          |      |  |      |
| A;Residues:  | 1-3010 <CHE>   |  |      |          |      |  |      |
| A;Cross-references:  | GB:M84754  |  |      |          |      |  |      |
| C;Superfamily:   | hepatitis C virus genome polyprotein   |  |      |          |      |  |      |
| C;Keywords:  | ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin                         |  |      |          |      |  |      |
| F;2-115/Product:   | capsid protein C #status predicted <CPC>   |  |      |          |      |  |      |
| F;116-191/Product:   | envelope protein M #status predicted <CPC>   |  |      |          |      |  |      |
| F;192-389/Product:   | major envelope protein E #status predicted <EPM>   |  |      |          |      |  |      |
| F;390-727/Product:   | nonstructural protein NS1 #status predicted <NS1>  |  |      |          |      |  |      |
| F;730-1006/Product:  | nonstructural protein NS2 #status predicted <NS2>  |  |      |          |      |  |      |
| F;1230-1237/Region:  | nucleotide-binding motif A (P-loop)  |  |      |          |      |  |      |
| F;1312-1317/Region:  | nucleotide-binding motif B (P-loop)  |  |      |          |      |  |      |
| F;1316-1319/Region:  | DEKH motif   |  |      |          |      |  |      |
| F;1616-1862/Product:   | nonstructural protein NS3 #status predicted <NS3>  |  |      |          |      |  |      |
| F;1663-2013/Product:   | nonstructural protein NS1 #status predicted <NS1>  |  |      |          |      |  |      |
| F;2014-3010/Product:   | nonstructural protein NS5 #status predicted <NS5>  |  |      |          |      |  |      |
| F;1616-1862/Product:   | nonstructural protein NS3 #status predicted <NS3>  |  |      |          |      |  |      |
| F;1916-209/Region:   | nucleotide-binding motif A (P-loop)  |  |      |          |      |  |      |
| F;1312-1317/Region:  | nucleotide-binding motif B (P-loop)  |  |      |          |      |  |      |
| Query Match  | 94.1%; Score 1870; DB 1; Length 3010;  |  |      |          |      |  |      |
| Best Local Similarity  | 94.1%; Score 1870; DB 1; Length 3010;  |  |      |          |      |  |      |
| Matches  | 349; Conservative 16; Mismatches 14; Indels 0; Gaps 0;   |  |      |          |      |  |      |
| Qy   | 2  | LITLSPYKVLLARLWQYLITVRAHQLWIPNVGGRDAILLTCAVPHLFID              | 61   | Db       | 828  | LITLSPYKVFLARLWQYLITVRAHQLWIPNVGGRDAILLTCAVPHLFID              | 887  |
| Db   | 828  | LITLSPYKVFLARLWQYLITVRAHQLWIPNVGGRDAILLTCAVPHLFID              | 887  | Qy       | 62   | ITKULLAIFGPMLVQAGITKPYFVRAQGLRACMLVKAGGHVQMAFMKAALTGT          | 121  |
| Qy   | 62   | ITKULLAIFGPMLVQAGITKPYFVRAQGLRACMLVKAGGHVQMAFMKAALTGT          | 121  | Db       | 888  | ITKULLAIFGPMLVQAGITKPYFVRAQGLRACMLVKAGGHVQMAFMKAALTGT          | 947  |
| Db   | 888  | ITKULLAIFGPMLVQAGITKPYFVRAQGLRACMLVKAGGHVQMAFMKAALTGT          | 947  | Qy       | 122  | YVVDHILPLODWAHAGLDRIDAVAVEPVIFSVDMEVKITTGADTAACGDIISGLPVSAARRG | 181  |
| Qy   | 122  | YVVDHILPLODWAHAGLDRIDAVAVEPVIFSVDMEVKITTGADTAACGDIISGLPVSAARRG | 181  | Db       | 948  | YVVDHILPLODWAHAGLDRIDAVAVEPVIFSVDMEVKITTGADTAACGDIISGLPVSAARRG | 1007 |
| Db   | 948  | YVVDHILPLODWAHAGLDRIDAVAVEPVIFSVDMEVKITTGADTAACGDIISGLPVSAARRG | 1007 | Qy       | 182  | REILGPGDFEGQWRLLAPITAYQSQRGLGICITSLGRDKNQVEGEVQVSTATQ          | 241  |
| Qy   | 182  | REILGPGDFEGQWRLLAPITAYQSQRGLGICITSLGRDKNQVEGEVQVSTATQ          | 241  | Db       | 1008 | REILGPGDFEGQWRLLAPITAYQSQRGLGICITSLGRDKNQVEGEVQVSTATQ          | 1067 |
| Db   | 1008   | REILGPGDFEGQWRLLAPITAYQSQRGLGICITSLGRDKNQVEGEVQVSTATQ          | 1067 | Qy       | 242  | SFLATCNGVWTVFAGSKTLAPKPKPTOMYINVDLVGQAPPGARSMTPCTGS            | 301  |
| Qy   | 242  | SFLATCNGVWTVFAGSKTLAPKPKPTOMYINVDLVGQAPPGARSMTPCTGS            | 301  | Db       | 1068 | SFLATCNGVWTVFAGSKTLAPKPKPTOMYINVDLVGQAPPGARSMTPCTGS            | 1127 |
| Qy   | 1068   | SFLATCNGVWTVFAGSKTLAPKPKPTOMYINVDLVGQAPPGARSMTPCTGS            | 1127 | Qy       | 302  | SDLYLVRHADVIPRRRGSRGSLSPRIVSYLKGSGGPLCPSGHAGVIFRAAVCTR         | 361  |
| Db   | 1068   | SFLATCNGVWTVFAGSKTLAPKPKPTOMYINVDLVGQAPPGARSMTPCTGS            | 1127 | Db       | 1128 | SDLYLVRHADVIPRRRGSRGSLSPRIVSYLKGSGGPLCPSGHAGVIFRAAVCTR         | 1187 |
| Qy   | 302  | SDLYLVRHADVIPRRRGSRGSLSPRIVSYLKGSGGPLCPSGHAGVIFRAAVCTR         | 361  | Qy       | 362  | GVAKAVDFIPESMETMR  | 380  |
| Db   | 1128   | SDLYLVRHADVIPRRRGSRGSLSPRIVSYLKGSGGPLCPSGHAGVIFRAAVCTR         | 1187 | Db       | 1188 | GVAKAVDFIPESMETMR  | 1206 |
| Qy   | 362  | GVAKAVDFIPESMETMR  | 380  | Db       | 1188 | GVAKAVDFIPESMETMR  | 1206 |
| Qy   | 362  | GVAKAVDFIPESMETMR  | 380  | RESULT 3 |      |  |      |
| Db   | 1188   | GVAKAVDFIPESMETMR  | 1206 | Qy       | 302  | SDLYLVRHADVIPRRRGSRGSLSPRIVSYLKGSGGPLCPSGHAGVIFRAAVCTR         | 361  |
| Qy   | 302  | SDLYLVRHADVIPRRRGSRGSLSPRIVSYLKGSGGPLCPSGHAGVIFRAAVCTR         | 361  | Db       | 1128 | SDLYLVRHADVIPRRRGSRGSLSPRIVSYLKGSGGPLCPSGHAGVIFRAAVCTR         | 1187 |
| Qy   | 362  | GVAKAVDFIPESMETMR  | 380  | Qy       | 362  | GVAKAVDFIPESMETMR  | 380  |
| Db   | 1128   | SDLYLVRHADVIPRRRGSRGSLSPRIVSYLKGSGGPLCPSGHAGVIFRAAVCTR         | 1187 | Db       | 1188 | GVAKAVDFIPESMETMR  | 1206 |
| Qy   | 362  | GVAKAVDFIPESMETMR  | 380  | Db       | 1188 | GVAKAVDFIPESMETMR  | 1206 |
| Qy   | 362  | GVAKAVDFIPESMETMR  | 380  | Qy       | 302  | SDLYLVRHADVIPRRRGSRGSLSPRIVSYLKGSGGPLCPSGHAGVIFRAAVCTR         | 361  |
| Db   | 1188   | GVAKAVDFIPESMETMR  | 1206 | Db       | 1128 | SDLYLVRHADVIPRRRGSRGSLSPRIVSYLKGSGGPLCPSGHAGVIFRAAVCTR         | 1187 |
| Qy   | 302  | SDLYLVRHADVIPRRRGSRGSLSPRIVSYLKGSGGPLCPSGHAGVIFRAAVCTR         | 361  | Qy       | 362  | GVAKAVDFIPESMETMR  | 380  |
| Db   | 1128   | SDLYLVRHADVIPRRRGSRGSLSPRIVSYLKGSGGPLCPSGHAGVIFRAAVCTR         | 1187 | Db       | 1188 | GVAKAVDFIPESMETMR  | 1206 |
| Qy   | 362  | GVAKAVDFIPESMETMR  | 380  | Db       | 1188 | GVAKAVDFIPESMETMR  | 1206 |
| Qy   | 362  | GVAKAVDFIPESMETMR  | 380  | Qy       | 302  | SDLYLVRHADVIPRRRGSRGSLSPRIVSYLKGSGGPLCPSGHAGVIFRAAVCTR         | 361  |
| Db   | 1188   | GVAKAVDFIPESMETMR  | 1206 | Db       | 1128 | SDLYLVRHADVIPRRRGSRGSLSPRIVSYLKGSGGPLCPSGHAGVIFRAAVCTR         | 1187 |
| Qy   | 302  | SDLYLVRHADVIPRRRGSRGSLSPRIVSYLKGSGGPLCPSGHAGVIFRAAVCTR         | 361  | Qy       | 362  | GVAKAVDFIPESMETMR  | 380  |
| Db   | 1128   | SDLYLVRHADVIPRRRGSRGSLSPRIVSYLKGSGGPLCPSGHAGVIFRAAVCTR         | 1187 | Db       | 1188 | GVAKAVDFIPESMETMR  | 1206 |
| Qy   | 362  | GVAKAVDFIPESMETMR  | 380  | Db       | 1188 | GVAKAVDFIPESMETMR  | 1206 |
| Qy   | 362  | GVAKAVDFIPESMETMR  | 380  | Qy       | 302  | SDLYLVRHADVIPRRRGSRGSLSPRIVSYLKGSGGPLCPSGHAGVIFRAAVCTR         | 361  |
| Db   | 1188   | GVAKAVDFIPESMETMR  | 1206 | Db       | 1128 | SDLYLVRHADVIPRRRGSRGSLSPRIVSYLKGSGGPLCPSGHAGVIFRAAVCTR         | 1187 |
| Qy   | 302  | SDLYLVRHADVIPRRRGSRGSLSPRIVSYLKGSGGPLCPSGHAGVIFRAAVCTR         | 361  | Qy       | 362  | GVAKAVDFIPESMETMR  | 380  |
| Db   | 1128   | SDLYLVRHADVIPRRRGSRGSLSPRIVSYLKGSGGPLCPSGHAGVIFRAAVCTR         | 1187 | Db       | 1188 | GVAKAVDFIPESMETMR  | 1206 |
| Qy   | 362  | GVAKAVDFIPESMETMR  | 380  | Db       | 1188 | GVAKAVDFIPESMETMR  | 1206 |
| Qy   | 362  | GVAKAVDFIPESMETMR  | 380  | Qy       | 302  | SDLYLVRHADVIPRRRGSRGSLSPRIVSYLKGSGGPLCPSGHAGVIFRAAVCTR         | 361  |
| Db   | 1188   | GVAKAVDFIPESMETMR  | 1206 | Db       | 1128 | SDLYLVRHADVIPRRRGSRGSLSPRIVSYLKGSGGPLCPSGHAGVIFRAAVCTR         | 1187 |
| Qy   | 302  | SDLYLVRHADVIPRRRGSRGSLSPRIVSYLKGSGGPLCPSGHAGVIFRAAVCTR         | 361  | Qy       | 362  | GVAKAVDFIPESMETMR  | 380  |
| Db   | 1128   | SDLYLVRHADVIPRRRGSRGSLSPRIVSYLKGSGGPLCPSGHAGVIFRAAVCTR         | 1187 | Db       | 1188 | GVAKAVDFIPESMETMR  | 1206 |
| Qy   | 362  | GVAKAVDFIPESMETMR  | 380  | Db       | 1188 | GVAKAVDFIPESMETMR  | 1206 |
| Qy   | 362  | GVAKAVDFIPESMETMR  | 380  | Qy       | 302  | SDLYLVRHADVIPRRRGSRGSLSPRIVSYLKGSGGPLCPSGHAGVIFRAAVCTR         | 361  |
| Db   | 1188   | GVAKAVDFIPESMETMR  | 1206 | Db       | 1128 | SDLYLVRHADVIPRRRGSRGSLSPRIVSYLKGSGGPLCPSGHAGVIFRAAVCTR         | 1187 |
| Qy   | 302  | SDLYLVRHADVIPRRRGSRGSLSPRIVSYLKGSGGPLCPSGHAGVIFRAAVCTR         | 361  | Qy       | 362  | GVAKAVDFIPESMETMR  | 380  |
| Db   | 1128   | SDLYLVRHADVIPRRRGSRGSLSPRIVSYLKGSGGPLCPSGHAGVIFRAAVCTR         | 1187 | Db       | 1188 | GVAKAVDFIPESMETMR  | 1206 |
| Qy   | 362  | GVAKAVDFIPESMETMR  | 380  | Db       | 1188 | GVAKAVDFIPESMETMR  | 1206 |
| Qy   | 362  | GVAKAVDFIPESMETMR  | 380  | Qy       | 302  | SDLYLVRHADVIPRRRGSRGSLSPRIVSYLKGSGGPLCPSGHAGVIFRAAVCTR         | 361  |
| Db   | 1188   | GVAKAVDFIPESMETMR  | 1206 | Db       | 1128 | SDLYLVRHADVIPRRRGSRGSLSPRIVSYLKGSGGPLCPSGHAGVIFRAAVCTR         | 1187 |
| Qy   | 302  | SDLYLVRHADVIPRRRGSRGSLSPRIVSYLKGSGGPLCPSGHAGVIFRAAVCTR         | 361  | Qy       | 362  | GVAKAVDFIPESMETMR  | 380  |
| Db   | 1128   | SDLYLVRHADVIPRRRGSRGSLSPRIVSYLKGSGGPLCPSGHAGVIFRAAVCTR         | 1187 | Db       | 1188 | GVAKAVDFIPESMETMR  | 1206 |
| Qy   | 362  | GVAKAVDFIPESMETMR  | 380  | Db       | 1188 | GVAKAVDFIPESMETMR  | 1206 |
| Qy   | 362  | GVAKAVDFIPESMETMR  | 380  | Qy       | 302  | SDLYLVRHADVIPRRRGSRGSLSPRIVSYLKGSGGPLCPSGHAGVIFRAAVCTR         | 361  |
| Db   | 1188   | GVAKAVDFIPESMETMR  | 1206 | Db       | 1128 | SDLYLVRHADVIPRRRGSRGSLSPRIVSYLKGSGGPLCPSGHAGVIFRAAVCTR         | 1187 |
| Qy   | 302  | SDLYLVRHADVIPRRRGSRGSLSPRIVSYLKGSGGPLCPSGHAGVIFRAAVCTR         | 361  | Qy       | 362  | GVAKAVDFIPESMETMR  | 380  |
| Db   | 1128   | SDLYLVRHADVIPRRRGSRGSLSPRIVSYLKGSGGPLCPSGHAGVIFRAAVCTR         | 1187 | Db       | 1188 | GVAKAVDFIPESMETMR  | 1206 |
| Qy   | 362  | GVAKAVDFIPESMETMR  | 380  | Db       | 1188 | GVAKAVDFIPESMETMR  | 1206 |
| Qy   | 362  | GVAKAVDFIPESMETMR  | 380  | Qy       | 302  | SDLYLVRHADVIPRRRGSRGSLSPRIVSYLKGSGGPLCPSGHAGVIFRAAVCTR         | 361  |
| Db   | 1188   | GVAKAVDFIPESMETMR  | 1206 | Db       | 1128 | SDLYLVRHADVIPRRRGSRGSLSPRIVSYLKGSGGPLCPSGHAGVIFRAAVCTR         | 1187 |
| Qy   | 302  | SDLYLVRHADVIPRRRGSRGSLSPRIVSYLKGSGGPLCPSGHAGVIFRAAVCTR         | 361  | Qy       | 362  | GVAKAVDFIPESMETMR  | 380  |
| Db   | 1128   | SDLYLVRHADVIPRRRGSRGSLSPRIVSYLKGSGGPLCPSGHAGVIFRAAVCTR         | 1187 | Db       | 1188 | GVAKAVDFIPESMETMR  | 1206 |
| Qy   | 362  | GVAKAVDFIPESMETMR  | 380  | Db       | 1188 | GVAKAVDFIPESMETMR  | 1206 |
| Qy   | 362  | GVAKAVDFIPESMETMR  | 380  | Qy       | 302  | SDLYLVRHADVIPRRRGSRGSLSPRIVSYLKGSGGPLCPSGHAGVIFRAAVCTR         | 361  |
| Db   | 1188   | GVAKAVDFIPESMETMR  | 1206 | Db       | 1128 | SDLYLVRHADVIPRRRGSRGSLSPRIVSYLKGSGGPLCPSGHAGVIFRAAVCTR         | 1187 |
| Qy   | 302  | SDLYLVRHADVIPRRRGSRGSLSPRIVSYLKGSGGPLCPSGHAGVIFRAAVCTR         | 361  | Qy       | 362  | GVAKAVDFIPESMETMR  | 380  |
| Db   | 1128   | SDLYLVRHADVIPRRRGSRGSLSPRIVSYLKGSGGPLCPSGHAGVIFRAAVCTR         | 1187 | Db       | 1188 | GVAKAVDFIPESMETMR  | 1206 |
| Qy   | 362  | GVAKAVDFIPESMETMR  | 380  | Db       | 1188 | GVAKAVDFIPESMETMR  | 1206 |
| Qy   | 362  | GVAKAVDFIPESMETMR  | 380  | Qy       | 302  | SDLYLVRHADVIPRRRGSRGSLSPRIVSYLKGSGGPLCPSGHAGVIFRAAVCTR         | 361  |
| Db   | 1188   | GVAKAVDFIPESMETMR  | 1206 | Db       | 1128 | SDLYLVRHADVIPRRRGSRGSLSPRIVSYLKGSGGPLCPSGHAGVIFRAAVCTR         | 1187 |
| Qy   | 302  | SDLYLVRHADVIPRRRGSRGSLSPRIVSYLKGSGGPLCPSGHAGVIFRAAVCTR         | 361  | Qy       | 362  | GVAKAVDFIPESMETMR  | 380  |
| Db   | 1128   | SDLYLVRHADVIPRRRGSRGSLSPRIVSYLKGSGGPLCPSGHAGVIFRAAVCTR         | 1187 | Db       | 1188 | GVAKAVDFIPESMETMR  | 1206 |
| Qy   | 362  | GVAKAVDFIPESMETMR  | 380  | Db       | 1188 | GVAKAVDFIPESMETMR  | 1206 |
| Qy   | 362  | GVAKAVDFIPESMETMR  | 380  | Qy       | 302  | SDLYLVRHADVIPRRRGSRGSLSPRIVSYLKGSGGPLCPSGHAGVIFRAAVCTR         | 361  |
| Db   | 1188   | GVAKAVDFIPESMETMR  | 1206 | Db       | 1128 | SDLYLVRHADVIPRRRGSRGSLSPRIVSYLKGSGGPLCPSGHAGVIFRAAVCTR         | 1187 |
| Qy   | 302  | SDLYLVRHADVIPRRRGSRGSLSPRIVSYLKGSGGPLCPSGHAGVIFRAAVCTR         | 361  | Qy       | 362  | GVAKAVDFIPESMETMR  | 380  |
| Db   | 1128   | SDLYLVRHADVIPRRRGSRGSLSPRIVSYLKGSGGPLCPSGHAGVIFRAAVCTR         | 1187 | Db       | 1188 | GVAKAVDFIPESMETMR  | 1206 |
| Qy   | 362  | GVAKAVDFIPESMETMR  | 380  | Db       | 1188 | GVAKAVDFIPESMETMR  | 1206 |
| Qy   | 362  | GVAKAVDFIPESMETMR  | 380  | Qy       | 302  | SDLYLVRHADVIPRRRGSRGSLSPRIVSYLKGSGGPLCPSGHAGVIFRAAVCTR         | 361  |
| Db   | 1188   | GVAKAVDFIPESMETMR  | 1206 | Db       | 1128 | SDLYLVRHADVIPRRRGSRGSLSPRIVSYLKGSGGPLCPSGHAGVIFRAAVCTR         | 1187 |
| Qy   | 302  | SDLYLVRHADVIPRRRGSRGSLSPRIVSYLKGSGGPLCPSGHAGVIFRAAVCTR         | 361  | Qy       | 362  | GVAKAVDFIPESMETMR  | 380  |
| Db   | 1128   | SDLYLVRHADVIPRRRGSRGSLSPRIVSYLKGSGGPLCPSGHAGVIFRAAVCTR</       |      |          |      |  |      |

|    |      |  |      |    |
|----|------|--|------|----|
| QY | 302  | SDLYLVRHADVIPRERRGDSRGSSLSPPRPSYVLKQSSGGPLCISGHAVGIFRAVCT    | 361  | Db |
| QY | 1128 | SDLYLVRHADVIPRERRGDSRGSSLSPPRPSYVLKQSSGGPLCISGHAVGIFRAVCT    | 1187 | Db |
| QY | 362  | GVAKAVDFIPVESMETMR   | 380  | Db |
| QY | 1188 | GVAKAVDFIPVESMETMR   | 1206 | Db |
| QY | 361  | GVAKAVDFIPVESMETMR   | 380  | Db |
| QY | 1187 | GVAKAVDFIPVESMETMR   | 1206 | Db |
| QY | 301  | SDLYLVRHADVIPRERRGDSRGSSLSPPRPSYVLKQSSGGPLCISGHAVGIFRAVCT    | 360  | Db |
| QY | 1127 | SDLYLVRHADVIPRERRGDSRGSSLSPPRPSYVLKQSSGGPLCISGHAVGIFRAVCT    | 1186 | Db |
| QY | 302  | SDLYLVRHADVIPRERRGDSRGSSLSPPRPSYVLKQSSGGPLCISGHAVGIFRAVCT    | 361  | Db |
| QY | 241  | QSLFLATCVNGCVWTVHAGSKTLAGPKGPINQMYTNVDOLVGQAPPGRASMTCTCG     | 300  | Db |
| QY | 1067 | QSLFLATCVNGCVWTVHAGSKTLAGPKGPINQMYTNVDOLVGQAPPGRASMTCTCG     | 1126 | Db |
| QY | 61   | DIRKULLATFGPMLNQLQAGITKIVYFVRAGLIRACMLVRKAAGGHHYQMAFMKLAALTG | 120  | Db |
| QY | 887  | DIRKULLATFGPMLNQLQAGITKIVYFVRAGLIRACMLVRKAAGGHHYQMAFMKLAALTG | 945  | Db |
| QY | 1    | ALLTSPYKVLLARLTIWLQQLTRVRAHLQWIPNPLVNRGERDAITLITCAVHPPELF    | 60   | Db |
| QY | 827  | ALLTSPYKVLLARLTIWLQQLTRVRAHLQWIPNPLVNRGERDAITLITCAVHPPELF    | 886  | Db |
| QY | 2    | LITSPYKVLLARLTIWLQQLTRVRAHLQWIPNPLVNRGERDAITLITCAVHPPELF     | 61   | Db |
| QY | 828  | LITSPYKVLLARLTIWLQQLTRVRAHLQWIPNPLVNRGERDAITLITCAVHPPELF     | 887  | Db |
| QY | 62   | ITKULLATFGPMLNQLQAGITKIVYFVRAGLIRACMLVRKAAGGHHYQMAFMKLAALTG  | 121  | Db |
| QY | 888  | ITKULLATFGPMLNQLQAGITKIVYFVRAGLIRACMLVRKAAGGHHYQMAFMKLAALTG  | 947  | Db |
| QY | 122  | YVYHLLTPQDWAHLGLDLAVAEVPEVIFSDMEKILITQDADTAACGDIISGLPVASRG   | 181  | Db |
| QY | 948  | YVYHLLTPQDWAHLGLDLAVAEVPEVIFSDMEKILITQDADTAACGDIISGLPVASRG   | 1007 | Db |
| QY | 182  | REIILGPADNFEGQWRLLAPITAYSQQTGGLGCGITSITGRDKNQVGEVQVSTATQ     | 241  | Db |
| QY | 1008 | KEILLGPADSLLEGGRGLRLAPITAYSQQTGGLGCGITSITGRDKNQVGEVQVSTATQ   | 1067 | Db |
| QY | 242  | SFLATCVNGCVWTVHAGSKTLAGPKGPINQMYTNVDOLVGQAPPGRASMTCTCG       | 301  | Db |
| QY | 1068 | SFLATCVNGCVWTVHAGSKTLAGPKGPINQMYTNVDOLVGQAPPGRASMTCTCG       | 1127 | Db |
| QY | 181  | GRELLIGGRDNFEGQWRLLAPITAYSQQTGGLGCGITSITGRDKNQVGEVQVSTAT     | 240  | Db |

|  |      |   |      |    |      |  |      |
|--|------|---|------|----|------|--|------|
| Db   | 1128 | SDLYLVTRHADVIFVRRRGDSRGSLISPRPVSYLKSSGGPLLCPPGHAVG-FRAAVCTR | 1187 | Qy | 243  | FLATCNGVGVNTVFHGAGSKTLAGPKPITQWYTNQDLVWGQAPPGARSMTPTCGSS   | 302  |
| Qy   | 362  | GVAKAVDFIVVESMETTMR   | 380  | Db | 1069 | FLATCNGVGVNTVFHGAGSKTLAGPKPITQWYTNQDLVWGQAPPGARSMTPTCGSS   | 1128 |
| Db   | 1188 | GVAKAVDFIVVESMETTMR   | 1206 | Qy | 303  | DLYLVTRHADVIFVRRRGDSRGSLISPRPVSYLKSSGGPLLCSSGHAVG-FRAAVCTR | 362  |
| RESULT 6   |      |   |      | Db | 1129 | DLYLVTRHADVIFVRRRGDSRGSLISPRPVSYLKSSGGPLLCPPGHAVG-FRAAVCTR | 1188 |
| Genome polyprotein - hepatitis C virus (strain HCV-1)  |      |   |      | Qy | 363  | VAKAVDFIPVESMETTMR   | 380  |
| N;Contains: capsid protein C; envelope Protein M; hepacivirin (EC 3.4.21.98) (nonstructural protein NS4a); nonstructural protein NS4b; nonstructural protein NS5   |      |   |      | Db | 1189 | VAKAVDFIPVENLETMR  | 1206 |
| C;Species: hepatitis C virus   |      |   |      | Qy |      |  |      |
| C;Date: 30-Sep-1992 #sequence revision 30-Sep-1992 #text_change 19-Jan-2001  |      |   |      | Db |      |  |      |
| C;Accession: A39166; P0403; PQ0034   |      |   |      | Qy |      |  |      |
| R;Choo, Q.L.; Richman, K.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.; Choi, J.; Gen. Virol., 73, 1131-1141, 1992                                    |      |   |      | Db |      |  |      |
| A;Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to e   |      |   |      | Qy |      |  |      |
| A;Reference number: PQ0393; MUID:91228871; PMID:1316939  |      |   |      | Db |      |  |      |
| A;Accession: E04043  |      |   |      | Qy |      |  |      |
| A;Cross-references: GB:M62321; NID:9329873; PIDN:AAA45676.1; PID:9329874   |      |   |      | Db |      |  |      |
| R;Choo, S.; McDonish, F.; Holmes, E.C.; Dow, B.; Peutheuer, J.F.; Follett, B.; Yap, P.I.   |      |   |      | Qy |      |  |      |
| J. Gen. Virol., 73, 1131-1141, 1992  |      |   |      | Db |      |  |      |
| A;Title: Molecular organization and diversity of the hepatitis C virus.  |      |   |      | Qy |      |  |      |
| A;Reference number: A39166; MUID:91112826; PMID:1848704  |      |   |      | Db |      |  |      |
| A;Accession: A39166  |      |   |      | Qy |      |  |      |
| A;Molecule type: RNA   |      |   |      | Db |      |  |      |
| A;Residues: 1-3011 <CHO>   |      |   |      | Qy |      |  |      |
| A;Cross-references: EMBL:D00831; NID:g21511; PIDN:BA00705.1; PID:9221512   |      |   |      | Db |      |  |      |
| A;Experimental source: isolates B-b17  |      |   |      | Qy |      |  |      |
| C;Superfamily: hepatitis C virus genome polyprotein  |      |   |      | Db |      |  |      |
| C;Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural protein; heparan sulfate   |      |   |      | Qy |      |  |      |
| F1-115/Product: capsid protein C #status predicted <CPC>   |      |   |      | Db |      |  |      |
| F116-191/Product: envelope protein M #status predicted <CPC>   |      |   |      | Qy |      |  |      |
| F192-389/Product: major envelope protein E #status predicted <MEB>   |      |   |      | Db |      |  |      |
| F390-729/Product: nonstructural protein NS1 #status predicted <NS1>  |      |   |      | Qy |      |  |      |
| F730-1006/Product: nonstructural protein NS2 #status predicted <NS2>   |      |   |      | Db |      |  |      |
| F1007-1615/Product: hepacivirin #status predicted <NS3>  |      |   |      | Qy |      |  |      |
| F1230-1237/Region: nucleotide-binding motif A (P-loop)   |      |   |      | Db |      |  |      |
| F1312-1317/Region: nucleotide-binding motif B  |      |   |      | Qy |      |  |      |
| F1316-1319/Region: nucleotide-binding motif B  |      |   |      | Db |      |  |      |
| F1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>  |      |   |      | Qy |      |  |      |
| F1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>  |      |   |      | Db |      |  |      |
| F2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>  |      |   |      | Qy |      |  |      |
| F196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2077,22  |      |   |      | Db |      |  |      |
| Query Match 86.3%; Score 1715; DB 1; Length 3011; Best Local Similarity 82.8%; Pred. No. 2, 8e-134; Matches 313; Conservative 32; Mismatches 33; Indels 0; Gaps 0; |      |   |      | Qy |      |  |      |
| Qy   | 3    | LITSPPYYKVLRLIWWIQLYLITRVEAHLQWIPPINVRGGRDAILITCAVHPPELFDI  | 62   | Db | 829  | LITSPPYYKVLRLIWWIQLYLITRVEAHLQWIPPINVRGGRDAILITCAVHPPELFDI | 888  |
| Db   | 829  | LITSPPYYKVLRLIWWIQLYLITRVEAHLQWIPPINVRGGRDAILITCAVHPPELFDI  | 888  | Qy | 3    | LITSPPYYKVLRLIWWIQLYLITRVEAHLQWIPPINVRGGRDAILITCAVHPPELFDI | 62   |
| Qy   | 63   | TKULLAATFGPMLIQASITKVKYFVRQGLTRACMVRKAAGHYOMAFMKLAALTGY     | 122  | Db | 829  | LITSPPYYKVLRLIWWIQLYLITRVEAHLQWIPPINVRGGRDAILITCAVHPPELFDI | 888  |
| Db   | 889  | TKULLAATFGPMLIQASITKVKYFVRQGLTRACMVRKAAGHYOMAFMKLAALTGY     | 948  | Qy | 63   | TKULLAATFGPMLIQASITKVKYFVRQGLTRACMVRKAAGHYOMAFMKLAALTGY    | 122  |
| Db   | 889  | TKULLAATFGPMLIQASITKVKYFVRQGLTRACMVRKAAGHYOMAFMKLAALTGY     | 948  | Db | 889  | TKULLAATFGPMLIQASITKVKYFVRQGLTRACMVRKAAGHYOMAFMKLAALTGY    | 948  |
| Qy   | 123  | YVDHLLTPQDWHAHGDRDLYAVEPVIPSDEKVIITWGAFTAAGDITISGLPVSAARR   | 182  | Qy | 123  | YVDHLLTPQDWHAHGDRDLYAVEPVIPSDEKVIITWGAFTAAGDITISGLPVSAARR  | 182  |
| Db   | 949  | VYHLLTPRDWAHNGLRDLAVAEVPTVFSQSMETKLITWGAFTAACDITINGLPVSARR  | 1008 | Db | 949  | VYHLLTPRDWAHNGLRDLAVAEVPTVFSQSMETKLITWGAFTAACDITINGLPVSARR | 1008 |
| Qy   | 183  | EILGPADNFEGQWRLAPITAYSQOTRGIGLICITSLGDKQNEVGWQVSTATQS       | 242  | Qy | 183  | EILGPADNFEGQWRLAPITAYSQOTRGIGLICITSLGDKQNEVGWQVSTATQS      | 242  |
| Db   | 1009 | EILGPADGMVSKGWRLLAPITAYAQOTRGIGLICITSLGDKQNEVGWQVSTATQS     | 1068 | Db | 1009 | EILGPADGMVSKGWRLLAPITAYAQOTRGIGLICITSLGDKQNEVGWQVSTATQS    | 1068 |
| Qy   | 243  | FLATCNGVGVNTVFHGAGSKTLAGPKPITQWYTNQDLVWGQAPPGARSMTPTCGSS    | 302  | Qy | 243  | FLATCNGVGVNTVFHGAGSKTLAGPKPITQWYTNQDLVWGQAPPGARSMTPTCGSS   | 302  |

|   |       |   |      |
|---|-------|---|------|
| Db  | 1069  | FLATCINGVCWTVHAGTRITASPKGPVQIYNTNDQDLYGWPAQGSRSLTPCTCGSS  | 1128 |
| Qy  | 303   | DLYLVRHADYIPVRRGDSGLSLSPPVSTIKGGSGGRPLCPSSHAVGIFRAAVCTRG  | 362  |
| Db  | 1129  | DLYLVRHADYIPVRRGDSRSSLSPRPISYLGKGGGPPICPAGHVGIFRAAVCTRG   | 1188 |
| Qy  | 363   | VAKAVDIPVESMETMR 380                                      |      |
| Db  | 1189  | VAKAVDIPVLETTNR 1206                                      |      |
| RESULT 8  |       |   |      |
| GNWVCH  |       |   |      |
| Name: polyprotein - hepatitis C virus (strain H)  |       |   |      |
| Genome: polyprotein - hepatitis C virus (strain H) (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5)  |       |   |      |
| Nr: Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5)  |       |   |      |
| C;Species: hepatitis C virus  |       |   |      |
| A;Note: host Homo sapiens (man)   |       |   |      |
| C;Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001   |       |   |      |
| C;Accession: A36814; A41546   |       |   |      |
| R;Inchause, G.; Zebdeee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.  |       |   |      |
| submitted to GenBank, July 1992   |       |   |      |
| A;Description: Genome structure of the human prototype strain H of hepatitis C virus: c   |       |   |      |
| A;Reference number: A36814  |       |   |      |
| A;Accession: A36814   |       |   |      |
| A;Molecule type: genomic RNA  |       |   |      |
| A;Residues: 1-3011 <INC>  |       |   |      |
| A;Cross-references: GB: M67463; NID: 9329737; PID: A3A45534.1; PID: 9329738   |       |   |      |
| R;Inchause, G.; Zebdeee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.  |       |   |      |
| Proc. Natl. Acad. Sci. U.S.A. 88, 10292-10296, 1991   |       |   |      |
| A;Title: Genomic structure of the human prototype strain H of hepatitis C virus: comparison with other hepatitis C viruses  |       |   |      |
| A;Reference number: A41546  |       |   |      |
| A;Contents: annotation  |       |   |      |
| A;Note: neither amino acid nor nucleotide sequence is given   |       |   |      |
| C;Superfamily: hepatitis C virus genome polyprotein   |       |   |      |
| C;Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural protein   |       |   |      |
| F;1-115/Product: capsid protein C #status predicted <CP>  |       |   |      |
| F;116-191/Product: envelope protein M #status predicted <EP>  |       |   |      |
| F;192-389/Product: major envelope protein E #status predicted <ME>  |       |   |      |
| F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>  |       |   |      |
| F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>   |       |   |      |
| F;1007-1615/Product: hepacivirin #status predicted <NS3>  |       |   |      |
| F;1230-1237/Region: nucleotide-binding motif A (P-loop)   |       |   |      |
| F;1311-1317/Region: nucleotide-binding motif B  |       |   |      |
| F;1311-1317/Region: nucleotide-binding motif A (P-loop)   |       |   |      |
| F;1516-1862/Product: nonstructural protein NS4a #status predicted <NS4A>  |       |   |      |
| F;1863-2013/Product: nonstructural protein NS4b #status predicted <NS4B>  |       |   |      |
| F;2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>  |       |   |      |
| F;196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2240,23  |       |   |      |
| Query Match   |       |   |      |
| Best Local Similarity 85.7%; Score 1703; DB 1; Length 3011;   |       |   |      |
| Best Local Similarity 82.3%; Score 2.8e-133; Pred. No. 2.8e-133; Indels 0; Gaps 0; Mismatches 311; Conservative 34; MisMatches 33; Indels 0; Gaps 0; Mismatches 263; Length 3011;保守性 56; 错配 33; 缺口 0; 空白 0; |       |   |      |
| Query   | 3     | LTLSPYYKKVLLRLWQQLTTRYAHQWIPRLPVNPGGRDAILLTCAVHPFLFD      | 61   |
| Db  | 829   | IIFTLPFGYKKVNLVWWLQFLTRVEAQLHVWVPPLNVRGRDAVILLTCVHPALVFDI | 888  |
| Qy  | 63    | TKLILAIQGPMVQIAGITKPYFVFRAGQIRACMLVKRAGGHYVOMAMKLAITGY    | 122  |
| Db  | 889   | TKLILAIQGPMVQIAGITKPYFVFRAGQIRACMLVKRAGGHYVOMAMKLAITGY    | 948  |
| Qy  | 123   | WYDHHTPLQDMAHAGLRLDAVAVPVIQSQTGIGCITSLGDRDNQVEGEQVSTATQ   | 242  |
| Db  | 949   | VYNHAPLRLMAHNGLRLDAVAVPVIQSPTMREKLTWTGADTAQCDINGLPLVSRQQ  | 1008 |
| Qy  | 183   | ERLIGPADNFEGQWRLAPITAYAQDTGIGCITSLGDRDNQVEGEQVSTATQ       | 182  |
| Db  | 1009  | ERLIGPADNFEGQWRLAPITAYAQDTGIGCITSLGDRDNQVEGEQVSTATQ       | 1068 |
| Qy  | 243   | FLATCINGVCWTVHAGSKTLAGKPTIQMTNTDQDLYGWQAPGARSMTPTCGS      | 302  |
| Db  | 1129  | ADLYLVRHADYIPVRRGDSRSSLSPRPISYLGKGGGPPICPAGHVGIFRAAVCTR   | 1188 |
| RESULT 9  |       |   |      |
| JCS620  |       |   |      |
| Query   |       |   |      |
| Name: polyprotein - hepatitis C virus (isolate EUH1480)   |       |   |      |
| Genome: polyprotein - hepatitis C virus (isolate EUH1480) (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5)   |       |   |      |
| C;Species: hepatitis C virus  |       |   |      |
| C;Accession: JCS620   |       |   |      |
| R;Chamberlain, R.W.; Adams, N.J.; Taylor, L.A.; Simmonds, P.; Elliott, R.M.   |       |   |      |
| Biochem. Biophys. Res. Commun. 236, 44-49, 1997   |       |   |      |
| A;Title: The complete coding sequence of hepatitis C virus genotype 5a, the predominant   |       |   |      |
| A;Reference number: JCS620; MUID: 97366593; PMID: 9223423   |       |   |      |
| A;Accession: JCS620   |       |   |      |
| A;Molecule type: mRNA   |       |   |      |
| A;Residues: 1-3014 <CEA>  |       |   |      |
| A;Cross-references: GB: Y13184  |       |   |      |
| A;Experimental source: genotype 5a, which predominates in South Africa  |       |   |      |
| A;Note: the translation of the nucleotide sequence is not complete in this paper  |       |   |      |
| C;Superfamily: hepatitis C virus genome polyprotein   |       |   |      |
| C;Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serine   |       |   |      |
| F;2-15/Product: capsid protein C #status predicted <CP>   |       |   |      |
| F;16-191/Product: envelope protein M #status predicted <EP>   |       |   |      |
| F;192-389/Product: major envelope protein E #status predicted <ME>  |       |   |      |
| F;384-408/Region: hypervariable #status predicted   |       |   |      |
| F;390-730/Product: nonstructural protein NS2 #status predicted <NS2>  |       |   |      |
| F;731-1007/Product: nonstructural protein NS2 #status predicted <NS3>   |       |   |      |
| F;1231-1238/Region: nucleotide-binding motif A (P-loop)   |       |   |      |
| F;1313-1318/Region: nucleotide-binding motif B  |       |   |      |
| F;1517-1863/Region: DEXR motif  |       |   |      |
| F;1617-1863/Region: nonstructural protein NS4a #status predicted <NS4A>   |       |   |      |
| F;1864-2014/Product: nonstructural protein NS4b #status predicted <NS4B>  |       |   |      |
| F;2015-3014/Product: nonstructural protein NS5 #status predicted <NS5>  |       |   |      |
| F;2210-2249/Region: interferon sensitivity determining #status predicted  |       |   |      |
| Query Match   | 74.6% | Score 1492; DB 1; Length 3014;                            |      |
| Best Local Similarity 69.4%; Pred. No. 7.7e-115; Gaps 0; Mismatches 263; Conservative 56; MisMatches 33; Indels 0; Gaps 0; Mismatches 263; Length 3011;保守性 56; 错配 33; 缺口 0; 空白 0;                           |       |   |      |
| Query   | 2     | LITLSPYYKKVLLRLWQQLTTRYAHQWIPRLPVNPGGRDAILLTCAVHPFLFD     | 61   |
| Db  | 829   | IIFTLPFGYKKVNLVWWLQFLTRVEAQLHVWVPPLNVRGRDAVILLTCVHPALVFDI | 888  |
| Qy  | 62    | TKLILAIQGPMVQIAGITKPYFVFRAGQIRACMLVKRAGGHYVOMAMKLAITGT    | 121  |
| Db  | 889   | TKLILAIQGPMVQIAGITKPYFVFRAGQIRACMLVKRAGGHYVOMAMKLAITGT    | 948  |
| Qy  | 122   | WYDHHTPLQDMAHAGLRLDAVAVPVIQSQTGIGCITSLGDRDNQVEGEQVSTATQ   | 241  |
| Db  | 949   | YIYDHAPMWDAAAGIRENTIVATPIVEFAMETKIVTWTGADTAQCDINGLPLVSRQQ | 1008 |
| Qy  | 182   | PELIGPADNFEGQWRLAPITAYAQDTGIGCITSLGDRDNQVEGEQVSTATQ       | 182  |
| Db  | 1009  | RELFIGPADDTKTSWGRLLAPITAYAQDTGIGCITSLGDRDNQVEGEQVSTATQ    | 1068 |
| Qy  | 242   | SFLATCINGVCWTVHAGSKTLAGKPTIQMTNTDQDLYGWQAPGARSMTPTCGS     | 301  |
| Db  | 1069  | TFLIGCINGVCWTVHAGSKTLAGKPTIQMTNTDQDLYGWQAPGARSMTPTCGS     | 1128 |
| Qy  | 302   | SDLYLVRHADYIPVRRGDSRSSLSPRPISYLGKGGGPPICPAGHVGIFRAAVCTR   | 361  |
| Db  | 1129  | ADLYLVRHADYIPVRRGDSRSSLSPRPISYLGKGGGPPICPAGHVGIFRAAVCTR   | 1188 |

QY 362 GVAKAVDFIPIVYESMETMR 380  
   |||:|||:|||:|||:|||:|||:|||:  
 Db 1189 GVAKALEFVPEVENLETTMR 1207

RESULT 10

JQ1303 genome polyprotein - hepatitis C virus (isolate HC-J6) N; Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5)

C-Species: hepatitis C virus C-Accession: A40250; PQ0397; PQ0559 R; Okamoto, H.; Kurai, K.; Okada, S.I.; Yamamoto, K.; Lizuka, H.; Tanaka, T.; Fukuda, S.; A; Accession: JQ1303 C-Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 19-Jan-2001

C-Accession: JQ1303 C-Title: Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from a human; Reference number: JQ1303; MUID:92044440; PMID:1658196 A; Residues: 1-1303 A; Molecule type: genomic RNA A; Residues: 1-1303 <OKA>

A; Cross-references: GB:D00944; NID:9221650; PIDN:BA000792.1; PID:9221651 A; Experimental source: isolate HC-J6 from a Japanese individual C; Superfamily: hepatitis C virus genome polyprotein C; Keywords: ATP; glycoprotein; hydrolase; P-loop; polyProtein; serine proteinase; transmembrane protein; capsid protein C #status predicted <CPC> F;116-191-/Product: envelope protein M #status predicted <EPM> F;116-191-/Product: major envelope protein E #status predicted <ME> F;90-733-/Product: nonstructural protein NS1 #status predicted <NS1> F;73-1010-/Product: nonstructural protein NS2 #status predicted <NS2> F;001-619-/Product: hepacivirin #status predicted <NS3> F;113-1321/Region: nucleotide-binding motif B

F;1320-1323/Region: DEXH motif F;1620-1866/Region: nonstructural protein NS4a #status predicted <NA4> F;1867-2017/Region: nonstructural protein NS4b #status predicted <NB4> F;2018-303-/Product: nonstructural protein NS5 #status predicted <NS5> F;196, 209, 234, 305, 325, 417, 423, 430, 448, 477, 534, 542, 558, 578, 627, 649, 1091, 1217, 1259, 2038, 28 Query Match 70.0%; Score 1391; DB 1; Length 3033; Best Local Similarity 66.0%; Pred. No. 3e-107; Matches 250; Conservative 55; Mismatches 74; Indels 0; Gaps 0;

Qy 2 LILTSPYYKKVLLARLWNLQLITRVEAHQWIPPLVNVRGGDAEILLTCAVHPFLIFD 61  
   |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:  
 Db 832 LFTLTPGPKLRLRFWMLCYLLTAAEMWQSWAPPQVQRGERDGITIWAIVFCPSWFD 891

QY 62 ITKULLATFGPLNLVQAGTTKIVYFVRAQGLIRACMLVRKAAGGHYVOMATMKLAALTG 121  
   |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:  
 Db 892 ITKULLAVLGPAVLKGALTIVYFVRAHALRMCTWVPHLAGGRVQWMLALGRWTG 951

QY 122 YVDLHPLQDWAPAGLIDLAVAVPPTFSDBVKLTWGAATAAGCDIISGLPSARRG 181  
   |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:  
 Db 952 YVDLHPLQDWAPAGLIDLAVAVPPTFSDBVKLTWGAATAAGCDIISGLPSARRG 1011

QY 182 RETLIGRADPNFEGQWMLAPATAYSSQTRGICLTSITGDRKDNQVEGVQVSTATO 241  
   |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:  
 Db 1012 REVLGPADGYTSRGWSLAPITAYAQOUTRGLIGITVSMITRKDTQAGBQIQTIVTO 1071

QY 1 ALLTSPYYKKVLLARLWNLQLITRVEAHQWIPPLVNVRGGDAEILLTCAVHPFLIF 60  
   |||:|||:|||:|||:|||:|||:|||:  
 Db 831 SIFTLTPAKLRLRFWMLCYLLTAAEMWQSWAPPQVQRGERDGITIWAIVFCPSWFD 890

QY 121 TYVDLHPLQDWAPAGLIDLAVAVPPTFSDBVKLTWGAATAAGCDIISGLPSARR 180  
   |||:|||:|||:|||:|||:|||:|||:  
 Db 951 TYVDLHPLQDWAPAGLIDLAVAVPPTFSDBVKLTWGAATAAGCDIISGLPSARR 1010  
   |||:|||:|||:|||:|||:|||:  
 Db 891 EVTKULLAVLGPAVLKGALTIVYFVRAHALRMCTWVPHLAGGRVQWMLALGRWTG 950

QY 302 SOLYLIVRHDATPRRGDSKSLSPRPSYIKSSSGGLCPSGHAVGIFRAVCER 361  
   |||:|||:|||:|||:|||:|||:  
 Db 1132 VDLYLIVTRNAVDPARRGRGDRKGLSPPRSTIKSSGGVLCPRGHAVGVFRAVCSR 1191

QY 362 GYAKASIDPPIVETLDIVR 380  
   |||:|||:|||:|||:  
 Db 1192 GYAKASIDPPIVETLDIVR 1210

RESULT 11

GNWWJB

QY 301 SSDYLIVRHDATPRRGDSRGSLSPPRPSYIKSSSGGLCPSGHAVGIFRAVCER 360  
   |||:|||:|||:|||:|||:|||:  
 QY 1071 QFLGTSISGVMLTVYHGAGNKTLAGKGPIQYMTWDQDLVGWQAPGARSMTCTCG 1130

RESULT 12

T08841 Polyprotein - dourcucouli hepatitis GB virus A  
 C;Species: dourcucouli hepatitis GB virus A  
 C;Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 17-Nov-2000  
 C;Accession: T08841  
 R;Erker, J.C.; Desai, S.M.; Leary, T.P.; Chalmers, M.L.; Montes, C.C.; Mushahwar, J.; Gen. Virol. 79, 41-45, 1998  
 A;Title: Genomic analysis of two GB virus A variants isolated from captive monkeys  
 A;Reference number: 216486; NID: 98120818; PMID: 9460920  
 A;Accession: T08841  
 A;Status: translated from GB/EMBL/DBJ  
 A;Molecule type: mRNA  
 A;Residues: 1-3005 <ERK>  
 A;Cross-references: EMBL:AF023425; NID:92828599; PIDN: AAC40502.1; PID:92828600  
 C;Superfamily: hepatitis C virus genome polyprotein  
 C;Keywords: polyprotein

Query Match 20.6%; Score 408.5; DB 2; Length 3005;  
 Best Local Similarity 31.3%; Pred. No. 2.9e-25; Matches 111; Conservative 57; Mismatches 154; Indels 33; Gaps 9;

| QY | 41   | RGGRD---AIIILTCAVAHPLIFELIDITKULAIIRGQPLMVQAGITKVYPVRAQGLRAC    | 96   |
|----|------|---|------|
| Db | 813  | RRGRDWRVTAAWVAGLIFERREVRVCSAALTA-----LAALIDSIVDILETL-ILTA       | 864  |
| QY | 97   | MVVRKA-----AGGHHYVOMAFVKLAALTGTYYVTDHLPLODWAHAGIRDIAWAVEPV      | 148  |
| Db | 865  | QPARRARLLSLTFFGDDADITRAFTRVRLERGVTFQHCCQVSXCAAAXLDGVALERPV      | 924  |
| QY | 149  | IFSDMEVKILTIGWAGDTAACDIISCPVSPSBRGRGILLG---PADMFEQGQWRLAPITAY   | 206  |
| Db | 925  | SVTARDCYTVIYDAAATLACGQRVECLPUPVVAARGDRDVLVGFPPSVRALPPGFVPTLPVV- | 983  |
| QY | 207  | SQQTRGELLGCITTSLGDRDNQVQESEBVQVSTATOSFLATCNGWVWTGHAGSSKTLAG     | 266  |
| Db | 984  | MQRGIGFFSVWTKTSMLGRDBREBEGSISIVWLTGSTRSMGTCVNGYNTTHGSNARTLAG    | 1043 |
| QY | 267  | PKGQTQMYTNVDQDGWQAPPGRASMTPTCGSSDLYLVTRHADVIPVRRGDSRGL          | 326  |
| Db | 1044 | PVGPNCRWMSPSDQDVAVYPLPGASCLEQKGTQSWCIRN--DGALCHGRSKLVEL         | 1101 |
| QY | 327  | LSPRVPVSYLKKGSSGGPILCPSGHAVGIFRAAVCTGY-----AKAVDFTIVES          | 374  |
| Db | 1102 | DLPTEISDFRGSSGSPILCDEGHVSGMM--VSULHHRKVTVQRYKPWELPKDS           | 1155 |

RESULT 13

T08839 polyprotein - marmoset hepatitis GB virus A  
 C;Species: marmoset hepatitis GB virus A  
 C;Date: 23-Jul-1999 #sequence\_revision 23-Jul-1999 #text\_change 17-Nov-2000  
 C;Accession: T08839  
 R;Erker, J.C.; Leary, T.P.; Chalmers, M.L.; Montes, C.C.; Mushahwar, J.; Gen. Virol. 79, 41-45, 1998  
 A;Title: Genomic analysis of two GB virus A variants isolated from captive monkeys  
 A;Reference number: Z16486; NID: 98120818; PMID: 9460920  
 A;Accession: T08839  
 A;Status: translated from GB/EMBL/DBJ  
 A;Molecule type: genomic RNA  
 A;Residues: 1-970 <ERK>  
 A;Cross-references: EMBL:AF023424; NID:92828597; PIDN: AAC40501.1; PID:92828598  
 C;Superfamily: hepatitis C virus genome polyprotein  
 C;Keywords: polyprotein

| Matches | Local Similarity  | Conservative                               | Best | Pred. | No.  | Mismatches | Indels | Gaps | 12% |
|---------|---|--|------|-------|------|------------|--------|------|-----|
| 103     | 28.9%   | 51   | W    | M     | 133  | 1          | 1      | 1    | 20% |
| 65      | LILAIFFGP-----L   | MVUQAGITKPVYVRAOGIIRACMLVRKAAGGHYVOMAFMKIA | 66   | I     | 116  |            |        |      |     |
| 816     | LVAANWFPPRETAACVCAVAFILGFPGFDFDVDTILEVLVLSSENVLVRLLARVIDSLVAGDKLA |  | 817  |       | 875  |            |        |      |     |
| 117     | ALTGTFTVYDILTPDQ-WAHAG-----IRDIAVAVEPVIFPSDMEVKITWGADTA           | 118  |      |       | 165  |            |        |      |     |
| 876     | T---TWWLEVKURKRKNCPFLYAHAGQTRRTAQLRQWGALEPVAVHEDCAMVRDARTL        | 877  |      |       | 932  |            |        |      |     |
| 166     | ACGDIISGLPVSARGRERTEILGPDNFEQGWRL-----LAPITAYSOOTRGGLGCIT         | 167  |      |       | 219  |            |        |      |     |
| 933     | SCGQSUVFGKPVVARRODEVLIGVLNW---WEIJPGFVPTAPVUHT-HHGKGFFGVKT        | 934  |      |       | 987  |            |        |      |     |
| 220     | SLTGRDKRNQVEGEVQVYSTATQSFALATCNGVCWTUFGAGSKTLAGPKGRITQMYNTVD      | 221  |      |       | 279  |            |        |      |     |
| 988     | SMTGWDBDETEHVGNVVVIGLTSITRSMGICVNGVMYTTYHGSNARTLAAQMGPVNRSWWAS    | 989  |      |       | 104  |            |        |      |     |
| 280     | ODLVGQAPPARSMPCTCGSSDLYLVTBHDV1PVRRGDSRSILLS-----                 | 281  |      |       | 328  |            |        |      |     |
| 1048    | DDAVAVPLPVGAKCLPCKCOPQOGWV-----RND-GALCHGHTGRTVFLDL               | 1049                                       |      |       | 109  |            |        |      |     |
| 329     | PRPVSVIKGSSGGPLCPSGHAVGFRAAVCTRG-----VAKADFIPYRMETT               | 330  |      |       | 378  |            |        |      |     |
| 1995    | DAELCDPRGSSQSPILDEGHAVGMJL-ISVHLRGSRVTVGIRYTKPWTLPREAITHT         | 1996                                       |      |       | 1150 |            |        |      |     |

Db 249 DENP-SFKVQIVGHSILGGGASLTYIIRHQKEFASATCFFAPGTPNLMINGESGRHFI 307  
 QY 272 TQMYINVQDQLV--GWQAPPGAASWTPCTCGSDLYLVTRHADVIP-VRRCDSRSIL 327  
 Db 308 TTING--SLVPFRSASSVDDLRSEVTSSWSNDLRLQVEHTRVLSVVRSATAIGSRL 365  
 QY 328 SPRPVSYLKQSSGGIILCP--SGHAVGIFRRAVCTRGYAKAVD 368  
 Db 366 PSIASAKAVKAGAGAILRPVSSGIQVAAPFLVNGC--GKIKCID 406

## RESULT 15

VHHWH2 structural protein 2 precursor - hepatitis E virus (strain Burma)

C;Species: hepatitis E virus

C;Date: 30-Sep-1992 #sequence\_revision 30-Sep-1992 #text\_change 23-Jul-1993

C;Accession: C40778

R.Tam, A.W.; Smith, M.M.; Guerra, M.E.; Huang, C.C.; Bradley, D.W.; Fry, K.E.; Reyes, G.

Virology 185, 120-131, 1991

A;Title: Hepatitis E virus (HEV): molecular cloning and sequencing of the full-length vi-

A;Reference number: A40778; MUID:92024067; PMID:1926770

A;Accession: C40778

A;Molecule type: genomic RNA

A;Cross-references: GB:W73218; NID:9330023; PID:AA245736.1; PID:9330026

A;Note: the authors translated the codon CGC for residue 2 as Ala

C;Superrfamily: hepatitis E virus structural protein 2

C;Keywords: structural protein

F;1-22/Domain: signal sequence #status predicted &lt;SIG&gt;

F;23-660/Product: structural protein 2 #status predicted &lt;SP2&gt;

Query Match 5.2%; Score 102.5; DB 1; Length 660;

Best Local Similarity 19.8%; Pred. No. 1.5; Matches 82; Conservative 52; Mismatches 130; Indels 151; Gaps 19;

QY 63 TKULL--AIFGPMLVLAG----ITKVPPYR--AQGLTRACMEVKRKGAGHVVQMA 111  
 Db 151 TNULVLYAAPLSPILPLQGNTNTHIMATEASVNAQYRVARATRYRPLVPAVGGAIS 210  
 QY 112 FMKLAALTGTYYTDHLTPLQDWAHAGLRLDAVAYEPVVFSPMEVKLTITWGADTAACGDI 171  
 Db 211 FWQQTTPPTPSV-----DNNSITSSTDVRLIVQPGIASLVI----- 246  
 QY 172 SGLPVSAARRGRETLIGPAD--NFEQOWRLLAPI-TAYSQCRGLI-----GCITSLG 223  
 Db 247 -----PSERLAHRNQGWRSTGVAAEEATSGLYMLCTIGSLNSYTN 290  
 QY 224 -----RDKNQVEGEVQVSTATOSFL 244  
 Db 291 TPYTGALGILDFALESPRNLTIPGNTNTRVRRYSSTARHRLRGGADGTABLTIA TREM 350  
 QY 245 A---TCUNGV-----CWTWPH-----GAG-----SKTAGPKG-PIT 272  
 Db 351 KOLYFSTSTNGVSEIGRGIALTFLNADTLLGQLPELISSAGGQFLPSRVUSANGEPTV 410  
 QY 273 QMTNTVNDQDLVSMQAPPGARTSMWTPCTCGSSDYLV--TRIADVTPVRRGDSRG-SLLS 328  
 Db 411 KLYTSENA---QQDKGIAAPHDILGESESWVIQYDQHEDRTPSPAPSRSRFSVLR 466  
 QY 329 PRPVSVYK-----GSGGPPCLCPSCSHAVGIFRRAVCTRGYAKAVD FIPV 372  
 Db 467 ANDVLWLSLTAEBYDOSTYGSSTGPVYV-SDSVTILVNVATGAQAVARSLLWTKV 519

|  |  |      |      |              |
|--|--|------|------|--------------|
| Copyright (c) 1993 - 2004 Compugen Ltd.  | GenCore version 5.1.6  |      |      |              |
| DM protein - protein search, using sw model  |  |      |      |              |
| Run on:  | May 6, 2004, 09:09:55 ; Search time 7.93319 Seconds  |      |      |              |
| Perfect score:   | US-10-650-585-12   |      |      |              |
| Sequence:  | 1 ALTISPVKVLARLWIL....., RGVAKAVDFTPVERSMETMR 380  |      |      |              |
| scoring table:   | BLOSUM62   |      |      |              |
| searched:  | Gapped 10.0 , Gapext 0.5   |      |      |              |
| total number of hits satisfying chosen parameters:   | 141681   |      |      |              |
| maximum DB seq length:   | 0  |      |      |              |
| post-processing:   | Minimum Match 0%<br>Maximum Match 100%<br>Listing first 45 summaries   |      |      |              |
| Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution. |  |      |      |              |
| SUMMARIES  |  |      |      |              |
| RESULT 1   |  |      |      |              |
| RP   | POLG_HCVJA   |      |      |              |
| RP   | POLG_HCVJA STANDARD; PRT; 3010 AA.   |      |      |              |
| RA   | P26650; ID POLG_HCVJA  |      |      |              |
| AC   | Rel. 23, Created)  |      |      |              |
| DT   | 01-AUG-1992 (Rel. 23, Last sequence update)  |      |      |              |
| DT   | 28-FEB-2003 (Rel. 41, last annotation update)  |      |      |              |
| DE   | Genome polyprotein [Contains: Capsid protein C (Core protein) (P22); Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)]  |      |      |              |
| DE   | (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivin)   |      |      |              |
| DE   | (EC 3.4.21.98); Nonstructural protein NS4A (P41); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48).  |      |      |              |
| DE   | Hepatitis C virus (isolate Japanese) (HCV). Hepatitis C viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus; OC  |      |      |              |
| DR   | NCHI_TaxID=11116;  |      |      |              |
| RN   | [1]  |      |      |              |
| SEQUENCE FROM N.A.   |  |      |      |              |
| RP   | MEDLINE=91088150; PubMed=2175903;  |      |      |              |
| RA   | Kato N., Hijikata M., Ootsuyama Y., Nakagawa M., Ohkoshi S., Sugimura T., Shimotohno K.;   |      |      |              |
| RT   | "Molecular cloning of the human hepatitis C virus genome from Japanese patients with non-A, non-B hepatitis";  |      |      |              |
| RL   | Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528(1990).  |      |      |              |
| RN   | [2]  |      |      |              |
| RP   | DISCUSSION OF SEQUENCE.  |      |      |              |
| RX   | MEDLINE=9119260; PubMed=1849488;   |      |      |              |
| RA   | Okkohi S., Shimotohno K., Nakagawa M., Ootsuyama Y., Muraiso K., RA  |      |      |              |
| RA   | "Molecular structure of the Japanese hepatitis C viral genome.";   |      |      |              |
| RL   | FBS Lett. 28:325-328(1991).  |      |      |              |
| CC   | -!- FUNCTION: The small proteins NS2A, NS3B, NS4A and NS4B are hydrophobic, suggesting a possible membrane-related function. NS3 and NS5 may play a role in the viral RNA replication.   |      |      |              |
| CC   | -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polyprotein, commonly with Asp or Glu in the P6 position. Cys or Thr in P1 and Ser or Ala in P1'.  |      |      |              |
| CC   | -!- CATALYTIC ACTIVITY: N nucleoside triphosphate + (RNA) (N).   |      |      |              |
| CC   | -!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a lipoprotein envelope. The envelope consists of two proteins: protein M and glycoprotein E. The nucleocapsid is a complex of protein C and mRNA.   |      |      |              |
| CC   | -!- PROTEIN C AND mRNA.  |      |      |              |
| CC   | -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.  |      |      |              |
| CC   | This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ). |      |      |              |
| CC   | EMBL: D90208; BAA4233.1; .   |      |      |              |
| ALIGNMENTS   |  |      |      |              |
| 34   | 85   | 4..3 | 3313 | 1 CLR3 RAT   |
| 35   | 84..5  | 4..3 | 1705 | 1 PTPV MOUSE |
| 35   | 84   | 4..2 | 309  | 1 UCP2 RAT   |
| 37   | 84   | 4..2 | 339  | 1 GPDA COREF |
| 38   | 84   | 4..2 | 403  | 1 MRPT ECOLI |
| 39   | 84   | 4..2 | 423  | 1 MK09 MOUSE |
| 40   | 84   | 4..2 | 470  | 1 NRAT TARUS |
| 41   | 84   | 4..2 | 503  | 1 YA02 MYCTU |
| 42   | 83..5  | 4..2 | 538  | 1 DAC ACTSP  |
| 43   | 83..5  | 4..2 | 1399 | 1 RPOC_PSEAB |
| 44   | 83   | 4..2 | 341  | 1 NDHM_BRANA |
| 45   | 82..5  | 4..2 | 453  | 1 NHAC_BACSU |
| 003744 brassica napus  |  |      |      |              |
| 007553 bacillus subtilis   |  |      |      |              |

|             |  | Best Local Similarity 94.2%; Pred. No. 6.9e-148; Matches 357; Conservative 12; Mismatches 10; Indels 0; Gaps 0; |  |
|-------------|--|---|--|
| DR          | PIR; A39053; GINVCU.   |   |  |
| DR          | HSSP; P20663; 1JXP.  |   |  |
| DR          | MEROPS; S29.001; -.  |   |  |
| DR          | MEROPS; 039.001; -.  |   |  |
| DR          | InterPro; IPR09003; Cys_Ser_trypsin.   |   |  |
| DR          | InterPro; IPR01101; DEAD.  |   |  |
| DR          | InterPro; IPR002522; HCV capsid.   |   |  |
| DR          | InterPro; IPR02521; HCV core.  |   |  |
| DR          | InterPro; IPR02519; HCV_env.   |   |  |
| DR          | InterPro; IPR02531; HCV_NS1.   |   |  |
| DR          | InterPro; IPR02518; HCV_NS2.   |   |  |
| DR          | InterPro; IPR00755; HCV_NS4a.  |   |  |
| DR          | InterPro; IPR01490; HCV_NS4b.  |   |  |
| DR          | InterPro; IPR02888; HCV_NS5a.  |   |  |
| DR          | InterPro; IPR02166; HCV_RdRp.  |   |  |
| DR          | InterPro; IPR01550; Helicase_C.  |   |  |
| DR          | InterPro; IPR04109; Peptidase_C19.   |   |  |
| DR          | InterPro; IPR07035; RNA_Pol_DS_ps.   |   |  |
| DR          | InterPro; IPR00794; RNA_Pol_PstI.  |   |  |
| DR          | Pfam; PF01543; HCV_capped.   |   |  |
| DR          | Pfam; PF01542; HCV_core.   |   |  |
| DR          | Pfam; PF01539; HCV_env.  |   |  |
| DR          | Pfam; PF01560; HCV_NS1.  |   |  |
| DR          | Pfam; PF01538; HCV_NS2.  |   |  |
| DR          | Pfam; PF02907; HCV_NS3.  |   |  |
| DR          | Pfam; PF01005; HCV_NS4a.   |   |  |
| DR          | Pfam; PF01001; HCV_NS4b.   |   |  |
| DR          | Pfam; PF01506; HCV_NS5a.   |   |  |
| DR          | Pfam; PRO0271; helicase_C.   |   |  |
| DR          | Protein; PR0098; Viral_RNP.  |   |  |
| DR          | SMART; SM00487; DEXDC_1.   |   |  |
| KW          | CORE_protein; Glycoprotein; Transferase; RNA-directed RNA polymerase;  |   |  |
| KW          | Coat_protein; Coat_protein; Envelope_protein; Helicase; ATP-binding;   |   |  |
| KW          | Transmembrane; Nonstructural protein; Hydrolase; Serine_protease;  |   |  |
| FT          | INIT_MER   | 1   |  |
| FT          | CHAIN  | 1   |  |
| FT          | CHAIN  | 115   |  |
| FT          | CHAIN  | 192   |  |
| FT          | CHAIN  | 384   |  |
| FT          | CHAIN  | 729   |  |
| FT          | CHAIN  | 730   |  |
| FT          | CHAIN  | 1007  |  |
| FT          | CHAIN  | 1615  |  |
| FT          | CHAIN  | 1862  |  |
| FT          | CHAIN  | 1863  |  |
| FT          | CHAIN  | 3010  |  |
| FT          | TRANSMEM   | 347   |  |
| FT          | ACT_SITE   | 1083  |  |
| FT          | ACT_SITE   | 1107  |  |
| FT          | ACT_SITE   | 1165  |  |
| FT          | NP_BIND  | 1230  |  |
| FT          | SITE   | 1316  |  |
| FT          | CARBONID   | 196   |  |
| FT          | CARBONID   | 209   |  |
| FT          | CARBONID   | 234   |  |
| FT          | CARBONID   | 250   |  |
| FT          | CARBONID   | 305   |  |
| FT          | CARBONID   | 417   |  |
| FT          | CARBONID   | 423   |  |
| FT          | CARBONID   | 430   |  |
| FT          | CARBONID   | 448   |  |
| FT          | CARBONID   | 532   |  |
| FT          | CARBONID   | 556   |  |
| FT          | CARBONID   | 576   |  |
| FT          | CARBONID   | 623   |  |
| FT          | CARBONID   | 645   |  |
| FT          | CARBONID   | 2041  |  |
| FT          | CARBONID   | 2077  |  |
| FT          | CARBONID   | 2240  |  |
| FT          | CARBONID   | 2788  |  |
| SO          | SEQUENCE   | 3010 AA:  |  |
| SO          | SEQUENCE   | 327017 MW:  |  |
| SO          | SEQUENCE   | AA991794 P6DB185 CRG64;   |  |
| SO          | SEQUENCE   | 327017 MW:  |  |
| SO          | SEQUENCE   | AA991794 P6DB185 CRG64;   |  |
| QY          | 2 LILTSYKVKULLARIWLMQLVITRVEAHQWIPPLNVRGRDILLLTCAVHPLFD  | 61  |  |
| Db          | 828 LILTSYKVKULLARIWLMQLVITRVEAHQWIPPLNVRGRDILLLTCAVHPLFD  | 887   |  |
| QY          | 62 ITKULLAIGFLMLQAGIKKVPVRAQSLIRACMLVKAAGHYQMAFMKLAALT   | 121   |  |
| Db          | 888 ITKULLAIGFLMLQAGIKKVPVRAQSLIRACMLVKAAGHYQMAFMKLAALT  | 947   |  |
| QY          | 122 YYDHILTPLODWAHAGLRDLAVAEVPUISDMEVKITTGADTAACGDIISGLPSA   | 181   |  |
| Db          | 948 YYDHILTPLODWAHAGLRDLAVAEVPUISDMETKLITWGADTRACGDIISGLPSA  | 1007  |  |
| QY          | 182 RBLIGGPADIFEGQWLRPITAVSQSQRGLGICISLTGRKQNEGEVQUSTAQ  | 241   |  |
| Db          | 1008 KEILLGPADSFGEQWMLRIPATIVSQSQTG  | 1067  |  |
| QY          | 302 SDLYUTRHDVLPFRRGDSRGALISPPRVSPVSYKGSGGPPUCPSGHAVGPFRAVCTR  | 361   |  |
| Db          | 1128 SDLYUTRHDVLPFRRGDSRGALISPPRVSYKGSGGPPUCPSGHAVGPFRAVCTR  | 1187  |  |
| QY          | 362 GVAKAVDFIVVESMTTR  | 380   |  |
| Db          | 1188 GVAKAVDFIVVESMTTR   | 1205  |  |
| RESULT 2    |  |   |  |
| ID          | POLG_HCVT  | STANDARD:   |  |
| AC          | Q00769;  | PRT; 3010 AA.   |  |
| DT          | 01-APR-1993 (Rel. 25, Created)   |   |  |
| DT          | 01-APR-1993 (Rel. 25, Last sequence update)  |   |  |
| DT          | 10-OCT-2003 (Rel. 42, Last annotation update)  |   |  |
| DE          | Genome polyprotein [Contains: Capsid protein C (Core protein) (P22), Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2 (GP88) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21) (EC 3.4.22.-); Protease/helicase (P70) (Hepacivirin) (EC 3.21.98); Nonstructural protein NS4A (P41); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48); Hepatitis C virus (isolate HC-JR) (HCV). Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus; NCBI_TaxID=31642; RN [1]; Sequence from N.A. MEDLINE:92205714; PubMed:1318627; Tanaka T., Kato N., Nakagawa M., Ootsuyama Y., Cho M.J., Nakazawa T., Hikikata M., Ishimura Y., Shimotohno K.; RT "Molecular cloning of hepatitis C virus genome from a single Japanese carrier: sequence variation within the same individual and among infected individuals;" RL Virus Res. 23:39-53(1992). |   |  |
| CC          | -!- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are hydrophobic, suggesting a possible membrane-related function. NS3 and NS5 may play a role in the viral RNA replication.   |   |  |
| CC          | -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polyprotein, commonly with Asp or Glu in the P6 position, Cys or Thr in P1 and Ser or Ala in P1.   |   |  |
| CC          | -!- CATALYTIC ACTIVITY: N nucleoside triphosphate + (RNA) (N).   |   |  |
| CC          | -!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a lipidprotein envelope. The envelope consists of two proteins: protein M and glycoprotein E. The nucleocapsid is a complex of protein C and mRNAs.   |   |  |
| CC          | -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.  |   |  |
| Query Match | 95.7%; Score 1902; DB 1; Length 3010;  |   |  |

Page  
(1)

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation of the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial purposes requires a license agreement (see [http://www.isb-sib.ch/announce\\_cc.html](http://www.isb-sib.ch/announce_cc.html)) or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

---

CC EMBL; D11168; BRA01943.1; --  
DR PIR; A45573; A45573.  
DR MEROS; S29\_001; --  
DR MEROS; US9\_001; --  
DR InterPro; IPR009003; Cys\_Ser\_trypsin.  
DR InterPro; IPR001410; DEAN.  
DR InterPro; IPR002522; HCV capsid.  
DR InterPro; IPR002521; HCV\_core.  
DR InterPro; IPR002519; HCV\_env.  
DR InterPro; IPR002531; HCV\_NSI.  
DR InterPro; IPR002518; HCV\_NS2.  
DR InterPro; IPR00745; HCV\_NS4.  
DR InterPro; IPR001490; HCV\_NS4b.  
DR InterPro; IPR002868; HCV\_NS5a.  
DR InterPro; IPR002166; HCV\_RdRP.  
DR InterPro; IPR001650; Helicase\_C.  
DR InterPro; IPR004109; Peptidase\_C29.  
DR InterPro; IPR007095; RNA-pol\_DS\_PS.  
DR InterPro; IPR007094; RNA-Pol\_PSVir.  
DR Pfam; PF002907; HCV\_NS3; 1.  
DR Pfam; PF001006; HCV\_NS4a; 1.  
DR Pfam; PF01001; HCV\_Ns4b; 1.  
DR Pfam; PF01506; HCV\_Ns5a; 1.  
DR Pfam; PF01533; HCV-env; 1.  
DR Pfam; PF01560; HCV\_NsI; 1.  
DR Pfam; PF01538; HCV\_NS2; 1.  
DR Pfam; PF01543; HCV\_capsid; 1.  
DR Pfam; PF000998; Viral\_RdRP; 1.  
DR Prodrom; PD186062; HCV\_NS1; 1.  
KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;  
Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;  
Transmembrane; Nonstructural protein; Hydrolase; Serine protease;  
Init-Met

FT 1 115 CAPSID PROTEIN C (POTENTIAL).  
FT 116 191 MATRIX PROTEIN (POTENTIAL).  
FT 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).  
FT 384 729 NONSTRUCTURAL PROTEIN NS1\_E2 (POTENTIAL).  
FT 730 1006 NON-STRUCTURAL PROTEIN NS2 (POTENTIAL).  
FT 1007 1615 PROTEASE/HELICASE NS3 (POTENTIAL).  
FT 1616 1862 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).  
FT 1863 2013 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).  
FT 2014 3019 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).  
FT 347 369 POTENTIAL.  
FT 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT 1230 1237 ATP (POTENTIAL).  
FT 1319 1319 DECH BOX.  
FT 196 196 N-LINKED (GLCNAC, . .) (POTENTIAL).  
FT 209 209 N-LINKED (GLCNAC, . .) (POTENTIAL).  
FT 234 234 N-LINKED (GLCNAC, . .) (POTENTIAL).  
FT 250 250 N-LINKED (GLCNAC, . .) (POTENTIAL).  
FT 305 305 N-LINKED (GLCNAC, . .) (POTENTIAL).  
FT 417 417 N-LINKED (GLCNAC, . .) (POTENTIAL).  
FT 423 423 N-LINKED (GLCNAC, . .) (POTENTIAL).  
FT 430 430 N-LINKED (GLCNAC, . .) (POTENTIAL).  
FT 448 448 N-LINKED (GLCNAC, . .) (POTENTIAL).  
FT 532 532 N-LINKED (GLCNAC, . .) (POTENTIAL).  
FT 540 540 N-LINKED (GLCNAC, . .) (POTENTIAL).  
FT 556 556 N-LINKED (GLCNAC, . .) (POTENTIAL).  
FT 576 576 N-LINKED (GLCNAC, . .) (POTENTIAL).

|                       |   |  |       |                                      |                    |
|-----------------------|---|--|-------|--------------------------------------|--------------------|
| FT                    | CARBOHYD  | 623  | 623   | N-LINKED (GLCNAC. . . ) (POTENTIAL). |                    |
| FT                    | CARBOHYD  | 645  | 645   | N-LINKED (GLCNAC. . . ) (POTENTIAL). |                    |
| FT                    | CARBOHYD  | 2041   | 2041  | N-LINKED (GLCNAC. . . ) (POTENTIAL). |                    |
| FT                    | CARBOHYD  | 2077   | 2077  | N-LINKED (GLCNAC. . . ) (POTENTIAL). |                    |
| FT                    | CARBOHYD  | 2240   | 2240  | N-LINKED (GLCNAC. . . ) (POTENTIAL). |                    |
| FT                    | CARBOHYD  | 2529   | 2529  | N-LINKED (GLCNAC. . . ) (POTENTIAL). |                    |
| FT                    | CARBOHYD  | 2788   | 2788  | N-LINKED (GLCNAC. . . ) (POTENTIAL). |                    |
| SEQUENCE              | 3010 AA;  | 326573   | MW:   | 941C7435D642BB                       | CRC64;             |
| Query Match           |   | 95.1%  | Score | 1890;                                | DB 1; length 3010; |
| Best Local Similarity |   | 94.2%  | Pred. | No. 5, 7e-47;                        |                    |
| Matches               | 37;   | Conservative   |       | 8; Mismatches                        | 14; Indels         |
|                       |   |  |       | 0;                                   | gaps 0;            |
| QY                    | 2   | LITLSPYKVQLARLILWMLQYLITRVEAHLOQWIPPLVNRGRDAILTCAHPELFID       | 61    |                                      |                    |
| Db                    | 828   | LITLSPHYKVFLARLILWMLQYLITRVEAHLCWVWPPLVNRGRDAILTCAHPELFID      | 887   |                                      |                    |
| QY                    | 62  | ITKLLAIFGPMILQAGITKIVYFVRQGLTRACMLYKAAGHHYOMAFNKLALTGT         | 121   |                                      |                    |
| Db                    | 888   | ITKLLAIFGPMILQAGITKIVYFVRQGLTRACMLYKAAGHHYOMAFNKLALTGT         | 947   |                                      |                    |
| QY                    | 122   | YVVDILTPILQDWAKAIGRLDAVAVEPVIFSDMEVKILITWGADTAACGDIISGIPVSARRG | 181   |                                      |                    |
| Db                    | 948   | YVVDILTP-QDWAKAIGRLDAVAVEPVIFSDMEVKILITWGADTAACGDIISGIPVSARRG  | 1007  |                                      |                    |
| Qy                    | 182   | REITLGPADNPEGQWRLAPITAYSQORTGLGCGCTSLTGRDKDNQEVEQVSVSTATQ      | 241   |                                      |                    |
| Db                    | 1008  | REITLGPADSIEEGQWRLAPITAYQORTGLGICIVSITGRDKDNQEVEQVSVSTATQ      | 1067  |                                      |                    |
| QY                    | 242   | SFLATCVNGCWTFHAGSKTIAAGKPKRPTQIYNTDQDGWQAOPGARGAMTPCTGS        | 301   |                                      |                    |
| Db                    | 1068  | SFLATCVNGCWTFHAGSKTIAAGKPKRPTQIYNTDQDGWQAOPGARGAMTPCTGS        | 1127  |                                      |                    |
| Qy                    | 302   | SDLVNTRHADVTRRERDSRSLSLRPVSVTKGSSGGPLCPSSHAVGFRAYCTR           | 361   |                                      |                    |
| Db                    | 1128  | SDLVNTRHADVTRRERDSRSLSLRPVSVTKGSSGGPLCPSSHAVGFRAYCTR           | 1187  |                                      |                    |
| QY                    | 362   | GVAKAVDFIPVESVETMWR  | 380   |                                      |                    |
| Db                    | 1188  | GVAKAVDFIPVESVETMWR  | 1205  |                                      |                    |
| RESULT 3              |   |  |       |                                      |                    |
| POLG_HCVTW            | ID_BOLG_HCVTW   | STANDARD;  | PRT;  | 3010 AA.                             |                    |
| AC                    | P29146;   |  |       |                                      |                    |
| DT                    | 01-APR-1993   | (Rel. 25, Created)   |       |                                      |                    |
| DT                    | 01-APR-1993   | (Rel. 25, Last sequence update)                                |       |                                      |                    |
| DT                    | 10-OCT-2003   | (Rel. 42, Last annotation update)                              |       |                                      |                    |
| DE                    | Genome polyprotein [Contains: Capsid protein C (Core protein) (P22); Envelope Glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21) (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirus) (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P66) (P70) (RNA-directed RNA Polymerase) (EC 2.7.7.48)]. |  |       |                                      |                    |
| DE                    | Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA Polymerase) (EC 2.7.7.48). Hepatitis C virus (isolate Taiwan) (HCV). Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.  |  |       |                                      |                    |
| OX                    | NCBI_TaxID=31445;   |  |       |                                      |                    |
| RN                    | [1]   |  |       |                                      |                    |
| RP                    | SEQUENCE FROM N.A.  |  |       |                                      |                    |
| RX                    | MEDLINE=22230206; PubMed=1314449;   |  |       |                                      |                    |
| RA                    | Chen P.J., Lin M.H., Tai K.F., Liu P.C., Lin C.J., Chen D.S.; The Taiwanese hepatitis C virus genome: sequence determination and mapping the 5' termini of viral genomic and antigenomic RNA.;"   |  |       |                                      |                    |
| RT                    | Virology 188:102-113(1992).   |  |       |                                      |                    |
| RL                    | - - FUNCTION: The small proteins NS2A, NS3B, NS4A and NS4B are hydrophobic, suggesting a possible membrane-related function. NS3 and NS5 may play a role in the viral RNA replication.  |  |       |                                      |                    |
| CC                    | - - CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polyprotein, commonly with Asp or Glu in the P6 position, Cys or Thr in P1 and Ser or Ala in P1'.   |  |       |                                      |                    |

-!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + (RNA)(N).

-!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a lipoprotein envelope. The envelope consists of two Proteins: protein M and glycoprotein E. The nucleocapsid is a complex of protein C and mRNAs.

-!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: M84754; -, NOT\_ANNOTATED\_CDS.

PIR: A420424; GNWVW.

PDB: 1N64; 25-FBB-03.

MEROPS: S26. 001; -.

INTERPRO: IPR009003; CYS\_Ser\_trypsin.

INTERPRO: IPR00522; HCV\_capsid.

INTERPRO: IPR002521; HCV\_core.

INTERPRO: IPR00251; HCV\_env.

INTERPRO: IPR00531; HCV\_NS1.

INTERPRO: IPR002518; HCV\_NS2.

INTERPRO: IPR00745; HCV\_NS4a.

INTERPRO: IPR01490; HCV\_NS4b.

INTERPRO: IPR005868; HCV\_NS5a.

INTERPRO: IPR002166; HCV\_RdRP.

INTERPRO: IPR01655; Helicase\_C.

INTERPRO: IPR004109; Peptidase\_C29.

INTERPRO: IPR007095; RNA\_Pol\_DS\_PS.

INTERPRO: IPR007094; RNA\_Pol\_DS\_Vil.

PFAM: PF01543; HCV\_capsid.

PFAM: PF01542; HCV\_core.

PFAM: PF01539; HCV\_env.

PFAM: PF01560; HCV\_NS1.

PFAM: PF01538; HCV\_NS2.

PFAM: PF01537; HCV\_NS3.

PFAM: PF01506; HCV\_NS4.

PFAM: PF00271; helicase\_C.

PFAM: PF00998; Viral\_RdRP.

ProDom: PD18602; HCV\_NS1.

SMART: SM00487; DEXC\_1.

Core\_Protein; Glycoprotein; Transferase; RNA-directed RNA polymerase; Transmembrane; Nonstructural protein; Hydrolase; Serine protease; 3D-structure.

INIT\_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE CELLULAR AMINOPEPTIDASE.

CORE PROTEIN (POTENTIAL).

MATRIX PROTEIN (POTENTIAL).

MAJOR ENVELOPE PROTEIN E (POTENTIAL).

NONSTRUCTURAL PROTEIN NS1/NS2 (POTENTIAL).

NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).

PROTEASE/HELICASE NS3 (POTENTIAL).

NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).

NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).

RNA-DIRECTED RNA POLYMERASE POTENTIAL.

CHARGE RELAY SYSTEM (BY SIMILARITY).

ACT SITE 107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).

ACT SITE 1155 1165 ATP (POTENTIAL).

NP BIND 1230 1237 DECH BOX.

SITE 1316 1319 N-LINKED (GLCNAC. . .) (POTENTIAL).

CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 2529 2529 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 2788 2788 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 3010 AA; 327047 MW; AAD26755CDP#215 CRC64; SQ SEQUENCE 3010 AA; 327047 MW; AAD26755CDP#215 CRC64;

Query Match 94.1%; Score 1870; DB 1; Length 3010;

Best Local Similarity 92.1%; Preid. No. 3e-145; Mismatches 14; Indels 0; Gaps Matches 349; Conservative 16; Mismatches 14; Indels 0; Gaps

Qy 2 LITLSPYYKVLLARLILWQLYLITRVEAHQWVTPPLNVRSGRDAILLTCAWHPELF 61

Db 828 LITLSPYYKVMLARLILWQLYTTRAEAHQWVTPPLNVRSGRDAILLTCAWPELF 868

Qy 62 ITKLLIAIFGPMVLYMLOQITKPYFVFAQGIRACMVLRVKAGGHVOMAFKLAALTGT 12

Db 888 ITKLLIAITGPMVLYMLOQITKPYFVFAQGIRACMVLRVKAGGHVOMAFKLAALTGT 948

Qy 122 YTDHDLTPLQDWAHAGRLDAVVAEPVPSMEVKILTGWADTAACDISSIPVSARRG 18

Db 948 YTDHDLTPLQDWAHAGRLDAVVAEPVPSMEVKILTGWADTAACDISSIPVSARRG 10

Qy 182 REILLGGRADNTBEGQGRLLAPTTAYSQTRGILGCITSLTGRDRKNOVEGEVQVSTATQ 242

Db 1008 REILLGGRADNTBEGQGRLLAPTTAYSQTRGILGCITSLTGRDRKNOVEGEVQVSTATQ 10

Qy 242 SPLATCINGVWTUFGASKTLAGPKGPITOMYNTNDQDVGQDAPPGARSMTPCGGS 300

Db 1068 SPLATCINGVWTUFGASKTLAGPKGPITOMYNTNDQDVGQDAPPGARSMTPCGGS 11

Qy 302 SDIYLVTRHADYTPVRRGDSRGLSLSPRPSYKLKASSGGPLCPSCAHGVFFRAVCTR 36

Db 1128 SDIYLVTRHADYTPVRRGDSRGLSLSPRPSYKLKASSGGPLCPSCAHGVFFRAVCTR 11

Qy 362 GVAKAVDFIPFESMETMR 380

Db 1188 GVAKAVDFIPFESMETMR 1206

RESULT 4

POLG\_HCVBK STANDARD; PRT: 3010 AA.

ID POLG\_HCVBK P2663;

DT 01-AUG-1992 (Rel. 23; Created)

DT 01-AUG-1992 (Rel. 23; Last sequence update)

DT 10-OCT-2003 (Rel. 42; Last annotation update)

DE Genome polyprotein [Contains: Capsid protein C (core protein) (P22); Envelope glycoprotein E1 (GP2); (GP35); Envelope glycoprotein E2 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21) (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirus) (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA Polymerase) (EC 2.7.7.48)].

DE Hepatitis C virus (isolate BK) (HCV).

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.

OC NCBI\_TaxID=11105;

[1]



FT STRAND 1108 1112  
 FT STRAND 1120 1120  
 FT STRAND 1122 1122  
 FT STRAND 1129 1133  
 FT TURN 1135 1136  
 FT STRAND 1139 1144  
 FT STRAND 1149 1157  
 FT HELIX 1158 1161  
 FT TURN 1162 1163  
 FT TURN 1165 1166  
 FT STRAND 1168 1171  
 FT TURN 1172 1174  
 FT STRAND 1175 1186  
 FT TURN 1187 1188  
 FT STRAND 1189 1197  
 FT HELIX 1198 1202  
 FT TURN 1203 1204  
 FT STRAND 1204 1208  
 SQ SEQUENCE 3010 AA; 327189 MW: F8422D5BCCFPFD9 CRC64;

Query Match 91.7%; Score 1823; DB 1; Length 3010;  
 Best Local Similarity 91.0%; Pred. No. 2.2e-141;  
 Matches 345; Conservative 12; Mismatches 22; Indels 0; Caps 0;  
 QY 2 LITLSPYKVLLARLWQLQFTRTAEHLQWIPPLNVRGRDAILLTCAVHFLFD 61  
 Db 828 LITLSPYKVLLARLWQLQFTRTAEHLQWIPPLNVRGRDAILLTCAVHFLFD 887

QY 62 ITKLIAIFGILMVLQAGITKPYFRAQGLTRACMLVLRKAGGHVQAMKMLAALTGT 121  
 Db 888 ITKLIAIFGILMVLQAGITKPYFRAQGLTRACMLVLRKAGGHVQAMKMLAALTGT 947

QY 122 YVVDHPTPLQWAHQASRDLAVAEVIFSVSMEVKITWGADTAQGDIISGLPYSARRG 181  
 Db 948 YVNHHTPLRWPRAQGDRDLAVAEVIFSVSMEVKITWGADTAQGDIISGLPYSARRG 1007

QY 182 REILGPADNEEGQGHLRLLATYSSQQTGRLGILTSLGRDKAQVEGVQVSTATQ 241  
 Db 1008 KEILGPADSELEGSRGRLLRLLATYSSQQTGRLGILTSLGRDKAQVEGVQVSTATQ 1067

QY 242 SFLATCNGVWVTHGGASKTMLAGRKPGTQMYNTNDQDILVGWQRPQGARSMTPTCGS 301  
 Db 1068 SFLATCNGVWVTHGGASKTMLAGRKPGTQMYNTNDQDILVGWQRPQGARSMTPTCGS 1127

QY 302 SDLYVTPRHADIVPVRGGSRSGLSPRVSYLGSSGSPRLCISGHAVGIFAACTR 361  
 Db 1128 SDLYVTPRHADIVPVRGGSRSGLSPRVSYLGSSGSPRLCISGHAVGIFAACTR 1187

QY 362 GVAKDVFIVPESMFTMR 380  
 Db 1188 GVAKDVFIVPESMFTMR 1206

---

**RESULT 5**  
**POLG\_HCV1**

|    |   |                                |      |          |
|----|---|--------------------------------|------|----------|
| ID | POLG_HCV1   | STANDARD;                      | PRT: | 3011 AA. |
| AC | P26664;   | 01-AUG-1992 (Rel. 23, Created) |      |          |
| DT | 01-AUG-1992 (Rel. 23, Last sequence update)   |                                |      |          |
| DT | 10-OCT-2003 (Rel. 42, Last annotation update)   |                                |      |          |
| DE | Genome polyprotein (Contains: Capsid Protein C (Core protein) (P22); Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2 (GP68) (GP70); Protein P7; Nonstructural protein NS2 (P21) (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepatitis C virus); (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66)) (RNA-directed RNA polymerase) (EC 2.7.7.48).) |                                |      |          |
| DE | Hepatitis C virus (isolate 1) (HCV); Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus; NCBI_TaxID=1104; [1]; RN  |                                |      |          |
| RP | SEQUENCE FROM N.A.  |                                |      |          |

RX MEDLINE=91172826; PubMed=1848744;  
 RA Choo O.-L.; Richman K.H.; Han J.-H.; Berger K.; Lee C.; Dong C.;  
 RA Gallegos C.; Coit D.; Medina-Selby A.; Barr P.J.; Weiner A.J.;  
 RA Bradley D.W.; Kuo G.; Houghton M.;  
 RT "Genetic organization and diversity of the hepatitis C virus";  
 RL Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).  
 CC -!- FUNCTION: The small proteins NS3A, NS2B, NS4A and NS4B are hydrophobic, suggesting a possible membrane-related function. NS3 and NS5 may play a role in the viral RNA replication.  
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor Polyprotein, commonly with Asp or Glu in the P6 position. Cys or Thr in P1 and Ser or Ala in P1',  
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + (RNA) (N).  
 CC -!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a lipoprotein envelope. The envelope consists of two proteins:  
 CC -> protein M and glycoprotein E. The nucleocapsid is a complex of protein M and capsid.  
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement. (See <http://www.isb-sib.ch/announce/> or send an email to license@sb-sib.ch).  
 CC DR EMBL; M62321; AAA45676.1; -.  
 DR PIR; A39166; GNWC3.  
 DR PDB; 1A1V; 16-FEB-99.  
 DR PDB; 1HE1; 25-NOV-98.  
 DR MEROPS; S29.001; -.  
 DR U39.001; -.  
 DR InterPro; IPR009003; Cys\_Ser\_trypsin.  
 DR InterPro; IPR001410; DEAD.  
 DR InterPro; IPR002522; HCV capsid.  
 DR InterPro; IPR002521; HCV core.  
 DR InterPro; IPR002519; HCV\_env.  
 DR InterPro; IPR002531; HCV\_NS1.  
 DR InterPro; IPR002518; HCV\_NS2.  
 DR InterPro; IPR00745; HCV\_NS4a.  
 DR InterPro; IPR001490; HCV\_NS4b.  
 DR InterPro; IPR02868; HCV\_NS5a.  
 DR InterPro; IPR0166; HCV\_RdRp.  
 DR InterPro; IPR001650; Helicase\_C.  
 DR InterPro; IPR004109; Peptidase\_C29.  
 DR InterPro; IPR007055; RNA\_polymerase.  
 DR InterPro; IPR007054; RNA\_poly\_Pol.  
 DR Pfam; PF01543; HCV\_cappa.  
 DR Pfam; PF01542; HCV\_core.  
 DR Pfam; PF01539; HCV\_env.  
 DR Pfam; PF01560; HCV\_NS1.  
 DR Pfam; PF01539; HCV\_NS1.  
 DR Pfam; PF01538; HCV\_NS2.  
 DR Pfam; PF02007; HCV\_NS3.  
 DR Pfam; PF01006; HCV\_NS4.  
 DR Pfam; PF01001; HCV\_NS4b.  
 DR Pfam; PF01056; HCV\_NS5a.  
 DR Pfam; PF00271; helicase\_C.  
 DR Pfam; PF00998; viral\_RdRP.  
 DR Pfam; PD18052; HCV\_NS1.  
 DR SMART; SM00479; DEXDC.  
 KW Core protein; Glycoprotein; Transferase; RNA-directed RNA polymerase; Polyprotein; Glycoprotein; Coat protein; Envelope protein; Helicase; ATP-binding; Transmembrane; Nonstructural protein; Hydrolase; Serine protease; 3D-structure.  
 FT INIT\_MET 1 1 REMOVED FROM CAPSD PROTEIN C BY THE CELULAR AMINOPEPTIDASE.  
 FT CHAIN 1 115 CASPID PROTEIN C (POTENTIAL).  
 FT CHAIN 116 131 MATRIX PROTEIN C (POTENTIAL).  
 FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).  
 FT CHAIN 384 729 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).  
 FT CHAIN 730 1008 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).

PT CHAIN 1007 1615 PROTEASE/HELICASE NS3 (POTENTIAL).  
 PT CHAIN 1616 1862 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).  
 PT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).  
 PT TRANSMEM 2014 3011 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).  
 PT ACT-SITE 347 369 POTENTIAL.  
 PT ACT-SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 PT ACT-SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 PT ACT-SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 PT NP-BIND 1230 1287 ATP (POTENTIAL).  
 FT SITE 1316 1319 DECH BOX.

FT CARBOHYD 196 196 N-LINKED (GCNAC. -) (POTENTIAL).  
 FT CARBOHYD 209 209 N-LINKED (GCNAC. -) (POTENTIAL).  
 FT CARBOHYD 234 234 N-LINKED (GCNAC. -) (POTENTIAL).  
 FT CARBOHYD 305 305 N-LINKED (GCNAC. -) (POTENTIAL).  
 FT CARBOHYD 417 417 N-LINKED (GCNAC. -) (POTENTIAL).  
 FT CARBOHYD 423 423 N-LINKED (GCNAC. -) (POTENTIAL).  
 FT CARBOHYD 430 430 N-LINKED (GCNAC. -) (POTENTIAL).  
 FT CARBOHYD 448 448 N-LINKED (GCNAC. -) (POTENTIAL).  
 FT CARBOHYD 476 476 N-LINKED (GCNAC. -) (POTENTIAL).  
 FT CARBOHYD 532 532 N-LINKED (GCNAC. -) (POTENTIAL).  
 FT CARBOHYD 540 540 N-LINKED (GCNAC. -) (POTENTIAL).  
 FT CARBOHYD 556 556 N-LINKED (GCNAC. -) (POTENTIAL).  
 FT CARBOHYD 576 576 N-LINKED (GCNAC. -) (POTENTIAL).  
 FT CARBOHYD 623 623 N-LINKED (GCNAC. -) (POTENTIAL).  
 FT CARBOHYD 645 645 N-LINKED (GCNAC. -) (POTENTIAL).  
 FT CARBOHYD 2041 2077 N-LINKED (GCNAC. -) (POTENTIAL).  
 FT CARBOHYD 2240 2240 N-LINKED (GCNAC. -) (POTENTIAL).  
 FT CARBOHYD 2364 2364 N-LINKED (GCNAC. -) (POTENTIAL).  
 FT CARBOHYD 2789 2789 N-LINKED (GCNAC. -) (POTENTIAL).  
 SQ SEQUENCE 3011 AA; 32197 MW; 65FBC9447FC5AF9 CRC64;

Query Match

Best Local Similarity

Score

86.3%

;

DB 1;

Length

3011;

Matches

313;

Conservative

32;

MisMatches

33;

Indels

0;

Gaps

0;

R1

R2

R3

R4

R5

R6

R7

R8

R9

R10

R11

R12

R13

R14

R15

R16

R17

R18

R19

R20

R21

R22

R23

R24

R25

R26

R27

R28

R29

R30

R31

R32

R33

R34

R35

R36

R37

R38

R39

R40

R41

R42

R43

R44

R45

R46

R47

R48

R49

R50

R51

R52

R53

R54

R55

R56

R57

R58

R59

R60

R61

R62

R63

R64

R65

R66

R67

R68

R69

R70

R71

R72

R73

R74

R75

R76

R77

R78

R79

R80

R81

R82

R83

R84

R85

R86

R87

R88

R89

R90

R91

R92

R93

R94

R95

R96

R97

R98

R99

R100

R101

R102

R103

R104

R105

R106

R107

R108

R109

R110

R111

R112

R113

R114

R115

R116

R117

R118

R119

R120

R121

R122

R123

R124

R125

R126

R127

R128

R129

R130

R131

R132

R133

R134

R135

R136

R137

R138

R139

R140

R141

R142

R143

R144

R145

R146

R147

R148

R149

R150

R151

R152

R153

R154

R155

R156

R157

R158

R159

R160

R161

R162

R163

R164

R165

R166

R167

R168

R169

R170

R171

R172

R173

R174

R175

R176

R177

R178

R179

R180

R181

R182

R183

R184

R185

R186

R187

R188

R189

R190

R191

R192

R193

R194

R195

R196

R197

R198

R199

R200

R201

R202

R203

R204

R205

R206

R207

R208

R209

R210

R211

R212

R213

R214

R215

R216

R217

R218

R219

R220

R221

R222

R223

R224

R225

R226

R227

R228

R229

R230

R231

R232

R233

R234

R235

R236

R237

R238

R239

R240

R241

R242

R243

R244

R245

R246

R247

R248

R249

R250

R251

R252

R253

R254

R255

R256

R257

R258

R259

R260

R261

R262

R263

R264

R265

R266

R267

R268

R269

R270

R271

R272

R273

R274

R275

R276

R277

R278

R279

R280

R281

R282

R283

R284

R285

R286

R287

R288

R289

R290

R291

R292

R293

R294

R295

R296

R297

R298

R299

R300

R301

R302

R303

R304

R305

R306

R307

R308

R309

R310

R311

R312

R313

R314

R315

R316

R317

R318

|          |   |   |        |                       |  |  |
|----------|---|---|--------|-----------------------|--|--|
| DR       | InterPro; IPR002521; HCV core.  | PT  | STRAND | 1291                  | 1295   |  |
| DR       | InterPro; IPR002331; HCV_N51.   | PT  | HELIX  | 1286                  | 1301   |  |
| DR       | InterPro; IPR002318; HCV_N52.   | PT  | TURN   | 1302                  | 1303   |  |
| DR       | InterPro; IPR00745; HCV_N54a.   | PT  | STRAND | 1312                  | 1316   |  |
| DR       | InterPro; IPR00490; HCV_N54b.   | PT  | TURN   | 1317                  | 1319   |  |
| DR       | InterPro; IPR002868; HCV_N55a.  | PT  | HELIX  | 1323                  | 1335   |  |
| DR       | InterPro; IPR002166; HCV_RNP.   | PT  | TURN   | 1336                  | 1340   |  |
| DR       | InterPro; IPR00150; Helicase_C.   | PT  | STRAND | 1343                  | 1347   |  |
| DR       | InterPro; IPR004109; Peptidase_C29.   | PT  | TURN   | 1352                  | 1353   |  |
| DR       | InterPro; IPR007095; RNA_poli_P5.   | PT  | STRAND | 1360                  | 1361   |  |
| DR       | InterPro; IPR00794; RNA_poli_P5Vir.   | PT  | STRAND | 1362                  | 1366   |  |
| PFam     | PF01543; HCV_capsid_1.  | PT  | STRAND | 1368                  | 1368   |  |
| PFam     | PF01542; HCV_core_1.  | PT  | STRAND | 1373                  | 1375   |  |
| PFam     | PF01539; HCV_env_1.   | PT  | TURN   | 1376                  | 1377   |  |
| PFam     | PF01560; HCV_N51_1.   | PT  | STRAND | 1378                  | 1380   |  |
| PFam     | PF01558; HCV_N52_1.   | PT  | HELIX  | 1382                  | 1385   |  |
| DR       | PFam; PRO2907; HCV_N53_1.   | PT  | STRAND | 1389                  | 1393   |  |
| DR       | PFam; PRO106; HCV_N54a_1.   | PT  | HELIX  | 1397                  | 1409   |  |
| DR       | PFam; PRO1001; HCV_N55b_1.  | PT  | TURN   | 1410                  | 1411   |  |
| DR       | PFam; PRO1506; HCV_N55a_1.  | PT  | STRAND | 1414                  | 1417   |  |
| DR       | PFam; PRO0271; helicase_C_1.  | PT  | TURN   | 1420                  | 1420   |  |
| DR       | Pratm; PRO098; viral_RdRP_1.  | PT  | STRAND | 1432                  | 1436   |  |
| DR       | Pratm; PRO86062; HCV_N51_1.   | PT  | TURN   | 1438                  | 1439   |  |
| DR       | SMART; SM00487; DEXDC_1.  | PT  | STRAND | 1450                  | 1453   |  |
| KW       | Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase; Core protein; Envelope protein; Helicase; ATP-binding; Transmembrane; Nonstructural protein; Hydrolase; Serine protease; 3D-structure. | PT  | STRAND | 1456                  | 1463   |  |
| INIT_MET | 1 1   | REMOVED FROM CAPSID PROTEIN C BY THE CELLULAR AMINOPEPTIDASE. | PT     | STRAND                | 1471   | 1478   |
| PT       | CHAIN 1 191   | CAPSID PROTEIN C.   | PT     | STRAND                | 1480   | 1480   |
| PT       | CHAIN 192 383   | ENVELOPE GLYCOPROTEIN E1.                                     | PT     | HELIX                 | 1481   | 1488   |
| PT       | CHAIN 384 746   | ENVELOPE GLYCOPROTEIN E2.                                     | PT     | HELIX                 | 1488   | 1488   |
| PT       | CHAIN 747 809   | PROTEIN_P.  | PT     | STRAND                | 1489   | 1490   |
| PT       | CHAIN 810 1026  | NONSTRUCTURAL PROTEIN NS2.                                    | PT     | HELIX                 | 1500   | 1501   |
| PT       | CHAIN 1027 1657   | PROTEASE/HELICASE_NS3.  | PT     | HELIX                 | 1507   | 1507   |
| PT       | CHAIN 1658 1711   | NONSTRUCTURAL PROTEIN NS4A.                                   | PT     | STRAND                | 1511   | 1511   |
| PT       | CHAIN 1712 1972   | NONSTRUCTURAL PROTEIN NS4B.                                   | PT     | HELIX                 | 1514   | 1527   |
| PT       | CHAIN 1973 2420   | NONSTRUCTURAL PROTEIN NS5A.                                   | PT     | HELIX                 | 1532   | 1544   |
| PT       | CHAIN 2421 3011   | NONSTRUCTURAL PROTEIN NS5B.                                   | PT     | STRAND                | 1550   | 1550   |
| PT       | TRANSMEM 347 369  | POTENTIAL.  | PT     | HELIX                 | 1555   | 1564   |
| PT       | ACT SITE 1083 1083  | CHARGE RELAY SYSTEM (BY SIMILARITY).                          | PT     | HELIX                 | 1570   | 1578   |
| PT       | ACT SITE 1107 1107  | CHARGE RELAY SYSTEM (BY SIMILARITY).                          | PT     | TURN                  | 1579   | 1580   |
| PT       | ACT SITE 1165 1165  | CHARGE RELAY SYSTEM (BY SIMILARITY).                          | PT     | HELIX                 | 1584   | 1597   |
| PT       | NP_BIND 1230 1237   | ATP (POTENTIAL).  | PT     | TURN                  | 1598   | 1598   |
| PT       | SITE 1316 1319  | DRCB BOX.   | PT     | HELIX                 | 1611   | 1611   |
| PT       | CARBOHYD 196 195  | (POTENTIAL).  | PT     | TURN                  | 1614   | 1618   |
| PT       | CARBOHYD 209 234  | N-LINKED (GLCNAC . .) (POTENTIAL).                            | PT     | STRAND                | 1622   | 1623   |
| PT       | CARBOHYD 234 234  | N-LINKED (GLCNAC . .) (POTENTIAL).                            | PT     | STRAND                | 1627   | 1627   |
| PT       | CARBOHYD 305 417  | N-LINKED (GLCNAC . .) (POTENTIAL).                            | PT     | STRAND                | 1635   | 1635   |
| PT       | CARBOHYD 417 423  | N-LINKED (GLCNAC . .) (POTENTIAL).                            | PT     | HELIX                 | 1640   | 1652   |
| PT       | CARBOHYD 423 430  | N-LINKED (GLCNAC . .) (POTENTIAL).                            | PT     | SEQUENCE              | 3011 AA:   | 327142 MW: 772CBB29CCD94753 CRC64;                             |
| PT       | CARBOHYD 430 448  | N-LINKED (GLCNAC . .) (POTENTIAL).                            | Query  | Match                 | 85.7%  | Score 1703; DB 1; length 3011;                                 |
| PT       | CARBOHYD 448 476  | N-LINKED (GLCNAC . .) (POTENTIAL).                            | Query  | Best Local Similarity | 82.3%  | Score 1703; DB 1; length 3011;                                 |
| PT       | CARBOHYD 476 532  | N-LINKED (GLCNAC . .) (POTENTIAL).                            | Db     | Matches               | 311; Conservative  | 34; Mismatches 33; Indels 0; Gaps 0;                           |
| PT       | CARBOHYD 532 540  | N-LINKED (GLCNAC . .) (POTENTIAL).                            | Db     | OY                    | 3  | LTTSPYYKTYLRLIWWQYLITRVEAMLQWTPPLNRGGGDAITLTCAVHPETLFFDI 62    |
| PT       | CARBOHYD 540 556  | N-LINKED (GLCNAC . .) (POTENTIAL).                            | Db     | 829                   | LTTSPPYYKRYISKWMWQYFLTRVEQLHVNWPPLNRRGGDAVILTCVHPALVFDI 888      |  |
| PT       | CARBOHYD 556 576  | N-LINKED (GLCNAC . .) (POTENTIAL).                            | Db     | OY                    | 63   | TKULLAIAQPMIYQAGTIVKPYFVRQGLIRACMULTRKAAGGHYVOMAFMKLAALTGY 122 |
| PT       | CARBOHYD 576 623  | N-LINKED (GLCNAC . .) (POTENTIAL).                            | Db     | 889                   | TKULLAIAQPLWLTIQASLLKVPYFVRQGLIRACMULTRKAAGGHYVOMAFMKLAALTGY 948 |  |
| PT       | CARBOHYD 623 645  | N-LINKED (GLCNAC . .) (POTENTIAL).                            | Db     | OY                    | 123  | WYHLTPPDWAFLAGRLAVAVEPVPSDMVKILTWTGAATTACCDISLCPVSARR 182      |
| PT       | CARBOHYD 645 1224   | (POTENTIAL).  | Db     | 949                   | VYNHLAPLRDWAHGLRLBLAVAEPVVPSRMETKLITWGADTAACGDTINGLPSARRQ 1008   |  |
| PT       | STRAND 1232 1233  | TURN  | QY     | 183                   | EILGPANPNEGQWRLAPITAYSQQTGQLLGCITSLTGRDNQNEGEVQNTQS 242          |  |
| PT       | TURN 1236 1238  |   | Db     | 1009                  | EILGPAGDGMVSKWRLAPITAYAQDTRGLGCLITSLTGRDNQEVGIVSTQT 1068         |  |
| PT       | TURN 1239 1246  |   | QY     | 243                   | FLATCVNSJCWTFPHGAGSKTAGPKFTQIYNTVODLVWQAPEGARNTPCGGS 302         |  |
| PT       | TURN 1247 1248  |   | Db     | 1069                  | FLATCINSVCWTHYAGGTRTASPCKPVIQYNTVODLVWPAQGSRLTPCCGSS 1128        |  |
| PT       | STRAND 1251 1255  |   | QY     | 303                   | DLYLVTRHADVIPVRRRDSRSGLSPRVSYLKGGGPLCPSHAGVIFRAAVCTRG 362        |  |
| PT       | TURN 1258 1271  |   | QY     |                       |  |  |
| PT       | TURN 1272 1277  |   | Db     |                       |  |  |
| PT       | STRAND 1277 1280  |   | QY     |                       |  |  |
| PT       | TURN 1281 1282  |   | Db     |                       |  |  |
| PT       | STRAND 1282 1283  |   | QY     |                       |  |  |

**RESULT 7**

ID POLG\_HCV6 STANDARD; PRT; 3033 AA.

AC P26660; HCV6

DT 01-AUG-1992 (Rel. 23, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P222); Envelope glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2 (GP68) (GP70); Protein P7; Nonstructural protein NS2 (P21) (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepatitis C virus); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA Polymerase) (EC 2.7.7.481); Hepatitis C virus ( isolate HC-6 ) (HCV); Hepatitis C viruses; sRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepaciviridae; NCBI\_TAXID=1113; RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=9204440; Pubmed=1658196;

RA Okamoto H., Okada S.-I., Sugiyama Y., Kurai K., Liuzuka H., Machida A., Miyakawa Y., Mayumi M.;

RT "Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from a human carrier: comparison with reported isolates for conserved RT and divergent regions"; J. Gen. Virol. 72:2697-2704(1991).

CC -!- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are hydrophobic, suggesting a possible membrane-related function. NS3 and NS5 may play a role in the viral RNA replication.

CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polyprotein commonly with Asp or Glu in the P6 position. Cys or Thr in P1 and Ser or Ala in P1'.

CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + RNA (N).

CC -!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a lipoprotein envelope. The envelope consists of two proteins: protein M and glycoprotein E. The nucleocapsid is a complex of protein C and mRNA.

CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; D00944; BAA0792.1; -.

DR PIR; JQ1303; JQ1303.

DR HSSP; P27958; IHER.

DR MEROPS; S29\_001; -.

DR InterPro; IPRO09003; Cys\_Ser\_trypsin.

DR InterPro; IPRO01410; DDA.

DR InterPro; IPRO02521; HCV\_Capsid.

DR InterPro; IPRO02521; HCV\_core.

DR InterPro; IPRO02519; HCV\_env.

DR InterPro; IPRO0531; HCV\_NS1.

DR InterPro; IPRO02518; HCV\_NS2.

DR InterPro; IPRO00745; HCV\_NS4a.

DR InterPro; IPRO01490; HCV\_NS4b.

DR InterPro; IPRO02868; HCV\_NS5a.

DR InterPro; IPRO02166; HCV\_RdRp.

DR InterPro; IPR001650; Helicase\_C.

DR InterPro; IPR004199; Peptidase\_C29.

DR InterPro; IP0007055; RNA\_poli\_DS\_ps.

DR InterPro; IPRO07054; RNA\_poli\_PSVR.

DR Pfam; PF01543; HCV\_cappid; 1.

DR Pfam; PF01542; HCV\_core; 1.

DR Pfam; PF01539; HCV\_en1; 1.

DR Pfam; PF01560; HCV\_NS1; 1.

DR Pfam; PF01538; HCV\_NS2; 1.

DR Pfam; PF02907; HCV\_NS3; 1.

DR Pfam; PF01006; HCV\_NS4a; 1.

DR Pfam; PF01001; HCV\_NS4b; 1.

DR Pfam; PF01506; HCV\_NS5a; 1.

DR Pfam; PF00271; helicase\_C; 1.

DR Pfam; PF00998; viral\_RdRp; 1.

DR SMART; SM00487; Dexx\_C; 1.

KW Polypeptide; Glycoprotein; Transferase; RNA-directed RNA Polymerase; Core protein; Coat protein; Envelope protein; Helicase; Amp-binding; Transmembrane; Nonstructural protein; Hydrolase; Serine protease.

FT INIT\_MET 1

FT CHAIN 1

FT CHAIN 115 115

FT CHAIN 116 191

FT CHAIN 192 383

FT CHAIN 384 733

FT CHAIN 734 1010

FT CHAIN 1011 1619

FT ACT SITE 1111

FT ACT SITE 1169 1169

FT ACT SITE 1234 1241

FT SITE 1320 1323

FT CARBOHYD 196 196

FT CARBOHYD 209 209

FT CARBOHYD 234 234

FT CARBOHYD 305 305

FT CARBOHYD 417 417

FT CARBOHYD 423 423

FT CARBOHYD 430 430

FT CARBOHYD 448 448

FT CARBOHYD 477 477

FT CARBOHYD 534 534

FT CARBOHYD 542 542

FT CARBOHYD 558 558

FT CARBOHYD 578 578

FT CARBOHYD 627 627

FT CARBOHYD 649 649

FT CARBOHYD 1091 1091

FT CARBOHYD 2038 2038

FT CARBOHYD 2811 2811

FT SEQUENCE 3033 AA; 329165 MW; F957F5C1A273BEE CRC34;

DR InterPro; IPR001650; Helicase\_C.

DR InterPro; IPR004199; Peptidase\_C29.

DR InterPro; IP0007055; RNA\_poli\_DS\_ps.

DR InterPro; IPRO07054; RNA\_poli\_PSVR.

DR Pfam; PF01543; HCV\_cappid; 1.

DR Pfam; PF01542; HCV\_core; 1.

DR Pfam; PF01539; HCV\_en1; 1.

DR Pfam; PF01560; HCV\_NS1; 1.

DR Pfam; PF01538; HCV\_NS2; 1.

DR Pfam; PF02907; HCV\_NS3; 1.

DR Pfam; PF01006; HCV\_NS4a; 1.

DR Pfam; PF01001; HCV\_NS4b; 1.

DR Pfam; PF01506; HCV\_NS5a; 1.

DR Pfam; PF00271; helicase\_C; 1.

DR Pfam; PF00998; viral\_RdRp; 1.

DR SMART; SM00487; Dexx\_C; 1.

KW Polypeptide; Glycoprotein; Transferase; RNA-directed RNA Polymerase; Core protein; Coat protein; Envelope protein; Helicase; Amp-binding; Transmembrane; Nonstructural protein; Hydrolase; Serine protease.

FT INIT\_MET 1

FT CHAIN 1

FT CHAIN 115 115

FT CHAIN 116 191

FT CHAIN 192 383

FT CHAIN 384 733

FT CHAIN 734 1010

FT ACT SITE 1111 1111

FT ACT SITE 1169 1169

FT ACT SITE 1234 1241

FT SITE 1320 1323

FT CARBOHYD 196 196

FT CARBOHYD 209 209

FT CARBOHYD 234 234

FT CARBOHYD 305 305

FT CARBOHYD 417 417

FT CARBOHYD 423 423

FT CARBOHYD 430 430

FT CARBOHYD 448 448

FT CARBOHYD 477 477

FT CARBOHYD 534 534

FT CARBOHYD 542 542

FT CARBOHYD 558 558

FT CARBOHYD 578 578

FT CARBOHYD 627 627

FT CARBOHYD 649 649

FT CARBOHYD 1091 1091

FT CARBOHYD 2038 2038

FT CARBOHYD 2811 2811

FT SEQUENCE 3033 AA; 329165 MW; F957F5C1A273BEE CRC34;

Query Match 70.0%; Score 1391; DB 1; Length 3033;

Best Local Similarity 66.0%; Pred. No. 7.1e-106; Gaps 0;

Matches 250; Conservative 55; Mismatches 74; Indels 0; Gaps 0;

QY 2 LITLSPVYKVLARLWLOVLTREVAHLWIPPLNVGRGRDAILTLCAVHRLBLFD 61

DB 832 LFTLTPGKYLISRFIWWLCYLTLIAEAQVWAPMQVRGRDGIIWAVAIIFCQVVF 891

QY 62 ITKLIAIFGLPMLQAGITKPVYFRAQGLIRACMVKRAGGGVQMAFMKLALST 121

DB 892 ITKWLIAVGPAYLKGATLPRVYFVRRAHALRCMVRHLAGGRVQWMLALGRWTG 951

QY 122 YVDHLLFLQDWAHAGJRLAVAVEVIFSDMEVKITWGAUTAACGDDISGLGVPASRG 181

DB 952 YIVDHLTMSDWAANGLRDLAVAVEPIIFSPMEKKVIVWGAETAACGDLINGLGPVSARLG 1011



|           |  |
|-----------|--|
| QY        | 61 DITKULIAATGFLMVQAGITKPVYVRAAGLIRACMLVRKAAGGHYVQAFMKIALTG 120<br>:::     :::   :::   :::   :::   :::   :::   :::   :::   :::   :::   :<br>Db 891 EVTWKWILALGPAVLKASLIRIPVYVRAHALRVCITLVKLAGARYIQMLITRGWTG 950<br>Db 121 TTYDHLSPLSLTAAGLIRDIAAVAVPVPVISDMEVKILITWGADTAACGDTISGLPVSSARR 180<br>QY 951 TTYDHLSPLSLTAAGLIRDIAAVAVPVPVISDMEVKILITWGADTAACGDTISGLPVSSARR 180<br>Db 181 GREILQPADFEGQSNRLAPITASQTSQRTGIGLQLGILTSLGRENQNEGEVOWSTAT 240<br>QY 1011 GREILQPADFEGQSNRLAPITASQTSQRTGIGLQLGILTSLGRENQNEGEVOWSTAT 240<br>Db 241 OSFLATCUNGVCWTFVHGAGSKTLAGPKGPITQMYTNDQDLYQHAPPGRSMTPCTCG 300<br>QY ::     :::   :::   :::   :::   :::   :::   :::   :::   :::   :::   :<br>Db 1071 QTFLGISISGVWTVHAGKNTLACPQKPVYQTMYSAGBLGVWNSPGTKSLIDPCTCG 1130<br>QY 301 SSDLYVTRHADVIPTRRGSRGSLSLSPRVSYLKQSGGAPLICPSGHAVGIFRAVCT 360<br>Db 1131 AVDLYVTRNADVIPVPRKDRRGALJSPPRISTLKQSSGAPVLCSSRGHAVGVLRAVCA 1190<br>QY 361 RGWAKAVDFPVESEMTEMR 380<br>Db 1191 REVAKSDDFPESLDWATR 1210  |
| RESULT 9  | VST2-HEVBU<br>ID_VST2-HEVBU STANDARD; PRT; 660 AA.<br>AC P29326;<br>DT 01-DEC-1992 (Rel. 24, Created)<br>DT 01-DEC-1992 (Rel. 24, Last sequence update)<br>DR 01-FEB-1994 (Rel. 28, Last annotation update)<br>DE Structural protein 2 precursor (ORF2).<br>OS Hepatitis E virus (strain Burman) (HEV).<br>OC Viruses; ssRNA positive-strand viruses, no DNA stage;<br>OC Hepatitis E-like viruses.<br>RN NCBI_TaxID=31767;<br>RP SEQUENCE FROM N.A.<br>RX MEDLINE=92024667;<br>RA Fry K.B., Reyes G.R., Guerra M.E., Huang C.-C., Bradley D.W., Tam A.W., Smith M.M.,<br>RT "Hepatitis E virus (HEV): molecular cloning and sequencing of the full-length viral genome";<br>RL VIROLOGY 185:120-131(1991).<br>CC -- FUNCTION: CONTAINS A HIGH BASIC AMINO ACID CONTENT SUGGESTING THAT IT MAY BE INVOLVED IN THE ENCAPSIDATION OF THE GENOMIC RNA BY EFFECTIVELY NEUTRALIZING THE NEGATIVELY CHARGED RNA.<br>CC -- This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to license@isb-sib.ch).                      |
| RESULT 10 | VST2-HEVPA<br>ID_VST2-HEVPA STANDARD; PRT; 660 AA.<br>AC P33426;<br>DT 01-FEB-1994 (Rel. 28, Created)<br>DT 01-FEB-1994 (Rel. 28, Last sequence update)<br>DR 01-FEB-1994 (Rel. 28, Last annotation update)<br>DE Structural protein 2 precursor (ORF2).<br>OS Hepatitis E virus (strain Pakistan) (HEV).<br>OC Viruses; ssRNA positive-strand viruses, no DNA stage;<br>OC Hepatitis E-like viruses.<br>RN NCBI_TaxID=33774;<br>RP SEQUENCE FROM N.A.<br>RX MEDLINE=92115700; PubMed=1731327;<br>RA Tsarev S.A., Emerson S.U., Reyes G.R., Tsareva T.S., Legters L.J., Malik I.A., Idai M., Purcell R.H., "Characterization of a prototype strain of hepatitis E virus";<br>RL Proc. Natl. Acad. Sci. U.S.A. 89:559-563 (1992).<br>CC -- FUNCTION: CONTAINS A HIGH BASIC AMINO ACID CONTENT SUGGESTING THAT IT MAY BE INVOLVED IN THE ENCAPSIDATION OF THE GENOMIC RNA BY EFFECTIVELY NEUTRALIZING THE NEGATIVELY CHARGED RNA.<br>CC -- This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to license@isb-sib.ch). |
| QY        | 151 TNLVLYAFLSPPLQGINTIMATERSANVQYVARATTRYRPLVNAVGGYASIS 210<br>:::     :::   :::   :::   :::   :::   :::   :::   :::   :::   :<br>Db 112 FMKLAALTCTYVTHLPLQDWAHGLRDAVAVPVPVISDMEVKILITWGADTAACGDLI 171<br>Db 211 FWPQTTPPTPSV-----DMMSITSDVRLVQPGIASELV----- 246<br>QY 172 SGIPUSARRGRELLGPAD--NFEQGWRILAPI-TAVSQQTGIL-----GCITSLTG 223<br>Db 247 -----PSBRLYRNQGRSVEETSGVAAEATSGVLMCLTHGSILVNSYN 290<br>QY 224 -----RDKQVEGVQVQVSTATQSFL 244<br>Db 291 TPVIGALGLDEALEFRNLTPGTNTVRVSSTARHRLRIGADGTAELTAAFRM 350<br>QY 245 A---TCVNGV----CWTVH-----GAG-----SKTLAGPKG PIT 272<br>Db 351 KDLFTISTINGVGEGRGIAITLFLADLILGGIPELSSAGQLEPYSRPVWSANGBTV 410<br>QY 273 QMYNTNDQDLYQHAPPGRSMTPCTCCSSLVIV--TRHADVIPVRRGDSRG-SLIS 328<br>Db 411 KLYTSEVNA---QODKGIAIPHDIDGESRSVYQDYQNCHEDRDPITSPARPPFSVLR 466<br>QY 329 PRPSYLK-----GSSGEPGQPSGHAVGIFRAVCTRGVAKAVDFP 372<br>Db 467 ANDVWLSLTAAEYDOSTYGSSITGPVVY--SDSVTLVNATGAQAVAKSDLWIKY 519  |

Db 151 TNLVVYAAPSPPLQDGTNTHMATEASNYAQVRATRYRPLPVNPGYALIS 210  
 Qy 112 FMKLAALTCTTYVHDHPLQDWAHAGLRLAVAVEPVFSDMEVKITWGADTAACDII 171  
 Db 211 FWPTTTPPTPSV-----DMNSITSTDVRLVOPGIAESLVI----- 246  
 Qy 172 SGLPVASARRGREELIGPAD---NEBGQWLL-----APITAYSQ 208  
 Db 247 -----PSSRLHTRNQGRSVEETSGVABEATSGLVMCLCSPNVSNT 290  
 Qy 209 QT-RGLGCIT-----ITSLGRDKNO-----VEGEVQVSTATOSFL 244  
 Db 291 TPYGALGLDFALELERNLITGCTINTRVRSSTARHRLRGADSTAEELTTAATRM 350  
 Qy 245 A---TCVNGV-----CWTVFH-----GAG----SKTAGPKG-PIT 272  
 Db 351 KDLVFTISTNGVETGRGIALTLFNADLTGLGIPTELISSAGQLFYSRPVVSANGEPTV 410  
 Qy 273 QMYNNVQDQIIVGQAPPGRSMPCCTCSSLVLUV--TRADVPIPRRDSRG-SLIS 328  
 Db 411 KLYVSVENA---QDGKGIAPIHDIDGRSRVVIQDVNQHQEDRPTSPSPSRPFSVLR 466  
 Qy 329 PRPSVYLK-----GSSCGPLCPSPGHAVGIRRAVTRGVAKAVFPIV 372  
 Db 467 ANDVILWLSLTAEVQDOSTYGSSTGPVY--SDSVTLVNVATGQAVERSLDWTKV 519  
 RESULT 11  
 SSBC\_ARATH STANDARD; PRT; 564 AA.  
 ID SSBC\_ARATH STANDARD; PRT; 564 AA.  
 AC P37107; 082570;  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Signal recognition particle 54 kDa protein, chloroplast precursor (SRP54) (54 chloroplast protein) (5ACP) (FFC).  
 GN FFC OR ATSG03940 OR FPF6 OR FPF5 150.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NOB1\_TAXID=3702;  
 RN [1]  
 RP STRAIN=NCV. Columbia;  
 RX MEDLINE=94012817; PubMed=8408079;  
 RA Franklin A.E., Hoffman N.E.;  
 RT "Characterization of a chloroplast homologue of the 54-kDa subunit of the signal recognition particle";  
 RL J. Biol. Chem. 268:22175-22180(1993).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RA Amin P., Sy D., Pilgrim M., Parry D.H., Hoffman N.E.;  
 RT "Isolation of two Arabidopsis mutants in the nuclear gene ffc, encoding the 54 kDa subunit of chloroplast signal recognition particle." (SEP-1998) to the EMBL/GenBank/DDBJ databases.  
 RL Submitted [3]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=NCV. Columbia;  
 RC MEDLINE=21016721; PubMed=11130714;  
 RX Tabata S., Kaneko T., Nakamura Y., Kotani H., Kato T., Asamizu E., Miyajima N., Sasamoto S., Kimura T., Hosouchi T., Kawashima K., Kohara M., Matsunoto M., Matsuno A., Muraki A., Nakayama S., Nakazaki N., Naruo K., Okumura S., Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M., Sato S., de la Bastide M., Huang E., Spiegel L., Gnoj L., O'Shaughnessy A., Preston R., Habermann K., Murray J., Johnson D., Rollifing T., Nelson J., Belter E., Pepin K., Spieth J., Sekhon M., Armstrong J., Becker M., Belter E., Cordum H., Cordeiro M., Courtney L., Dante M., Du H., Edwards J., Fryman J., Haakensen B., Lamar E., Latreille P., Leonard S., Meyer R., Mulvaney E., Ozersky P., Riley A., Strowmatt C., RA

Ra Wagner-McPherson C., Wollam A., Yoakum M., Bell M., Dednia N., Ra Parneil L., Shah R., Rodriguez M., Hoorn See L., Vil D., Baker J., Ra Kirchoff K., Toth K., King L., Bahret A., Miller B., Marra M.A., Ra Martenssen R., McCombie W.R., Wilson R.K., Murphy G., Bancroft I., Ra Volchert G., Wambatt R., Duesterhoft A., Stikema W., Pohl T., Ra Entian K.-D., Terry N., Hartley N., Bent E., Johnson S., Ra Langham S.-A., McCullagh B., Robben J., Grymonpre B., Zimmermann W., Ra Ramsperger U., Wedler H., Peters S., Ra van Staveren M., Birke W., Moeljiman P., Klein Lankhorst R., Ra Weitzenseger T., Bothe G., Rose M., Hauf J., Bernmeiser S., Feldpausch M., Lamberth S., Villarreal R., Gieben J., Arribalzaga W., Ra Bents O., Lemcke K., Kolesov G., Mayer K., Rudd S., Schoof H., Ra Schubert C., Zaccaria P., Mewes H.-W., Bevan M., Fransz P., Ra "Sequence and analysis of chromosome 5 of the plant Arabidopsis thaliana," Ra Nature 408:823-826 (2000).  
 CC FUNCTION: May target chloroplast proteins to either the thylakoid or envelope membranes.  
 CC -!- SUBCELLULAR LOCATION: Chloroplast stroma.  
 CC -!- TISSUE SPECIFICITY: Most abundant in green shoot tissue and lower levels seen in the roots and etiolated buds.  
 CC -!- SIMILARITY: Belongs to the Grp-binding SRP family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation in the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).  
 CC EMBL; 221970; CAV9981.1;  
 DR EMBL; AF092168; AAC64139.1;  
 DR EMBL; ALI62873; CBB85514.1; -.  
 DR PIR; S36637; S36637.  
 DR HSSP; 007347; 1FFH.  
 DR InterPro; IPR00353; AAA ATPase.  
 DR InterPro; IPR00877; SRP54.  
 DR InterPro; IPR004125; SRP54\_SPB.  
 DR InterPro; IPR004780; SRP\_sub.  
 DR Pfam; PF00448; SRP54; 1.  
 DR Pfam; PF02881; SRP54\_N; 1.  
 DR Pfam; PF02978; SRP\_SF\_B; 1.  
 DR ProDom; PD000819; SRP54; 1.  
 DR SMART; SM0322; AAA; 1.  
 DR TIGRFAMS; TIGR0059; ffh; 1.  
 DR PROSITE; PS0030; SRP54\_1.  
 DR Signal\_recognition\_particle; GRP-binding; RNA-binding; Chloroplast; KW Transit peptide.  
 FT TRANSIT 1 75 CHLOROPLAST.  
 FT CHAIN 76 564 SIGNAL RECOGNITION PARTICLE 54 kDa  
 FT DOMAIN 76 370 PROTEIN.  
 FT DOMAIN 371 564 G-DOMAIN.  
 FT NP BIND 183 190 GTP (BY SIMILARITY).  
 FT NP BIND 265 269 GTP (BY SIMILARITY).  
 FT NP BIND 323 326 GTP (BY SIMILARITY).  
 FT CONFLICT 76 76 E -> V (IN RBF. 2).  
 FT SEQUENCE 564 AA; 61232 MW; 423FT285FB9063E4 CRC64;  
 Qy Query Match 5.1%; Score 101; DB 1; Length 564;  
 BT Best Local Similarity 26.1%; Pred. No. 1;  
 BT Matches 54; Conservative 37; Mismatches 74; Indels 42; Gaps 11;  
 Qy 54 VHPPLI---FDITKLIAIFGSPLMVQAGI-----TKVYPYFVRAQGLTRACMLVR 100  
 DB 154 VHBLVLMGGEVSELQPAKSGPVTILLAGLQVGKTKTCALKKG-Q-KSCMLI- 210  
 Qy 101 KAAGGHHYQMAFKLAA---IGTYVVDHPLQ--DWAHAGLIRDIAWAVEPVFSDMEV 155  
 DB 211 --AGDVYPAIDOLVIGEIQVGVPTAGTIVPKADIKQGLKEAKR----MNVD 261  
 Qy 156 KITWGADTAACCDIISGLPVSAARRGREIL---IGPADNFGQQWRILLAPITAVSQ 210

|             |   |                   |   |
|-------------|---|-------------------|---|
| Db          | 262 VML---DTAGLQIDKQMDPLKDVCKFLNPTTEVILVVDAWTGQ--EAAALVTFENVEI 315  | QY                | 185 LIGPADNFECQWRLLAPITAYSQTRGILGCITSLGRDKNQVEGVQVSTATOSFL 244  |
| QY          | 211 RGLIGCITSLGRDKNQVEGVQVVS 237  | Db                | 145 LLSPPS-----FSPATPSQK-----YTSRNR----GEWVTFGSAQ--- 178  |
| Db          | -GITGAILTKLQDSRGGAALSVKEVs 341  | QY                | 245 ATCVNGVNTVFRGAGSKTL--AGPKGPIQTMVNVDQDVLG 284  |
| DT          | 01-FEB-1994 (Rel. 28, Created)  | Db                | 179 -----GLSWSGRGGGSVSLKVVSDPPEPTGSYKAMFOQLMG 215   |
| DT          | 01-FEB-1994 (Rel. 28, Last sequence update)   | RESULT 12         | RESULTS   |
| DE          | DNA polymerase alpha 70 kDa subunit (DNA polymerase subunit B).   | DP02_MOUSE        | DP02_MOUSE  |
| GN          | POLA2.  | ID                | STANDARD;   |
| OS          | Mus musculus (Mouse).   | PRT;              | 600 AA.   |
| OC          | Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurgnathi; Muridae; Murinae; Mus; OX  | NCBI_TaxId=10090; | NCBI_TaxId=10090;   |
| RX          | [1]   | RN                | SEQUENCE FROM N.A., AND SEQUENCE OF 84-102; 269-285 AND 394-403.  |
| RA          | Miyazawa H., Izumi M., Tada S., Takada R., Masutani M., Ui M., Hanaka F.;   | RT                | MEDLINE=93216788; PubMed=863324;  |
| RA          | Hanaka F.;  | RL                | PMID=93216788; PubMed=863324;   |
| RT          | Molecular cloning of the cDNAs for the four subunits of mouse DNA polymerase alpha-prime complex and their gene expression during cell proliferation and the cell cycle.;   | RT                | J. Biol. Chem. 268:8111-8122(1993).   |
| CC          | -!- FUNCTION: May play an essential role at the early stage of chromosomal DNA replication by coupling the polymerase alpha/prime complex to the cellular replication machinery. (By similarity).   | CC                | -!- SUBUNIT: DNA polymerase alpha-prime is a four subunit enzyme (subunits A, B, C and D), which is assembled throughout the cell cycle. The largest subunit (subunit A) has DNA polymerase activity, the two smallest subunits (subunits C and D) have DNA primase activity. Subunit B binds to subunit A.   |
| CC          | -!- SUBCELLULAR LOCATION: Nuclear.  | CC                | -!- SUBCELLULAR LOCATION: IN A CELL CYCLE-DEPENDENT MANNER, IN G2/M   |
| CC          | -!- PHASE (BY SIMILARITY):  | CC                | -!- PHASE (BY SIMILARITY):  |
| CC          | -!- SIMILARITY: Belongs to the DNA polymerase alpha subunit B family.   | CC                | -!- SIMILARITY: Belongs to the DNA polymerase alpha-prime complex.  |
| CC          | This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/or send an email to license@isb-sib.ch">http://www.isb-sib.ch/announce/or send an email to license@isb-sib.ch</a> ). | CC                | This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/or send an email to license@isb-sib.ch">http://www.isb-sib.ch/announce/or send an email to license@isb-sib.ch</a> ). |
| CC          | EMBL; D13546; BAA02746.1; -.  | CC                | EMBL; D13546; BAA02746.1; -.  |
| DR          | PIR; BA6642; BA6642.  | CC                | EMBL; U17042; AAA7998.1; -.   |
| DR          | MGD; MGI:93690; PolA.   | DR                | PIR; T18309; T18309.  |
| DR          | InterPro; IPR007200; DNA_pol_alpha_B.   | DR                | InterPro; IPR001054; G_cyclase.   |
| DR          | PFAM; PF04058; DNA_Pol_alpha_B_1.   | DR                | InterPro; PF00211; guanylate_cyc_1.   |
| FT          | DNA replication; Nuclear protein; Phosphorylation.  | DR                | SMART; SM0044; CYCC_1.  |
| FT          | DOMAIN 101 107 POLY-GLU   | KW                | PROSITE; PS50125; GUANYLATE_CYCLASES_2; 1.  |
| FT          | DOMAIN 115 157 PROSER/THR-RICH (HYDROPHILIC).   | FT                | Lyase; CAMP biosynthesis; Transmembrane; Receptor; Glycoprotein; Metal-binding; Magnesium.  |
| SQ          | SEQUENCE 600 AA; 66267 MN; 79F94BB6F33FEB3 CRC64;   | FT                | METAL-BINDING; MAGNESIUM (BY SIMILARITY).   |
| Query Match | Best Local Similarity 5.1%; Score 101; DB 1; Length 600; Matches 55; Conservative 34; Mismatches 71; Indels 62; Gaps 12;  | FT                | MAGNESIUM (BY SIMILARITY); N-LINKED (GLCNAC. . . ) (POTENTIAL); N-LINKED (GLCNAC. . . ) (POTENTIAL).   |
| QY          | 92 LIRACMLVLRKAAGHYYM-AFMKLAALT-----GTVVYDH-----TPQDWA 134  | FT                | CARBOHYD  |
| Db          | 27 LAELCVLYRQTEDGMVSELAFCAGKCLTWDILNSPEYEVINKKISKAKHSASDSC 86   | FT                | CARBONID  |
| QY          | 135 HAGRLDLAVAEVPIVFSDEMKVLTWGAFTACCDI-ISGLP-----VSARRGEI 184   | FT                | CARBONID  |
| Db          | 87 HAGRLDLAVAEVPIVFSDEMKVLTWGAFTACCDI-ISGLP-----VSARRGEI 184  | FT                | CARBONID  |
| QY          | 42 GGRDAILLTCAVHP---ELIFDITKL-----LIAIFGLM-----V 75   | FT                | SEQUENCE 1380 AA; 151692 MN; 6B2D5F7DC1107A0 CRC64;   |
| QY          | Query Match Best Local Similarity 21.0%; Pred. No. 13; Matches 93; Conservative 51; Mismatches 139; Indels 159; Gaps 24;  | QY                | Query Match Best Local Similarity 21.0%; Pred. No. 13; Matches 93; Conservative 51; Mismatches 139; Indels 159; Gaps 24;  |







GenCore version 5.1.6  
 Copyright (c) 1993 - 2004 Compugen Ltd.

### OM protein - protein search, using sw model

Run on:

May 6, 2004, 09:21:36 ; Search time 35.5407 Seconds

(Without alignments) 3773.509 Million cell updates/sec

Title: US-10-650-585-12

Perfect score: 1987

Sequence: 1 ALULSPVYKULLARIWL.....RGVAKAVDFIPVESMETMR 380

Scoring table: BLOSUM62

Gapext 0.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Maximum Match 0%

Listing first 45 summaries

Database : SPTREMBL\_25:\*

- 1: SP\_archea:\*
- 2: SP\_bacteria:\*
- 3: SP\_fungi:\*
- 4: SP\_human:\*
- 5: SP\_invertebrate:\*
- 6: SP\_mammal:\*
- 7: SP\_mbcl:\*
- 8: SP\_organelle:\*
- 9: SP\_phage:\*
- 10: SP\_plant:\*
- 11: SP\_rabbit:\*
- 12: SP\_virus:\*
- 13: SP\_vertebrate:\*
- 14: SP\_unclassified:\*
- 15: SP\_virus:\*
- 16: SP\_bacteriop:\*
- 17: SP\_archeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

### SUMMARIES

%

Result No. Score Query Length DB ID Description

| Result No. | Score | Query | Length | DB | ID     | Description        |
|------------|-------|-------|--------|----|--------|--------------------|
| 1          | 1921  | 96.7  | 3010   | 12 | Q9J3F9 | 09j3f9 hepatitis c |
| 2          | 1918  | 96.5  | 3010   | 12 | Q9DIE6 | 09d6e hepatitis c  |
| 3          | 1915  | 96.4  | 3010   | 12 | Q93H7  | 093h7 hepatitis c  |
| 4          | 1909  | 96.1  | 3010   | 12 | Q9DT6  | 09dt6 hepatitis c  |
| 5          | 1908  | 96.0  | 3010   | 12 | Q9DE4  | 09de4 hepatitis c  |
| 6          | 1907  | 96.0  | 3010   | 12 | Q80TP3 | 080tp3 hepatitis c |
| 7          | 1907  | 95.9  | 3010   | 12 | Q9J3G6 | 09j3g6 hepatitis c |
| 8          | 1905  | 95.9  | 3010   | 12 | Q68788 | 068788 hepatitis c |
| 9          | 1905  | 95.9  | 3010   | 12 | Q9J3H3 | 09j3h3 hepatitis c |
| 10         | 1905  | 95.9  | 3010   | 12 | Q9DPD7 | 09dpd7 hepatitis c |
| 11         | 1903  | 95.8  | 1186   | 12 | Q81755 | 081755 hepatitis c |
| 12         | 1903  | 95.8  | 284    | 12 | Q81817 | 081817 hepatitis c |
| 13         | 1903  | 95.8  | 3010   | 12 | P89966 | P89966 hepatitis c |
| 14         | 1902  | 95.7  | 3010   | 12 | Q9J3H9 | 09j3h9 hepatitis c |
| 15         | 1902  | 95.7  | 3010   | 12 | Q9GAU2 | 09gau2 hepatitis c |
| 16         | 1902  | 95.7  | 3010   | 12 | Q9J3I0 | 09j3i0 hepatitis c |

| RESULT 1 |  |      |          |  |  |  |
|----------|--|------|----------|--|--|--|
| Q9J3F9   | PRELIMINARY;   | PRT; | 3010 AA. |  |  |  |
| AC       | Q9J3F9;  |      |          |  |  |  |
| DT       | 01-OCT-2000 (TREMBrel. 15, Created)  |      |          |  |  |  |
| DT       | 01-OCT-2000 (TREMBrel. 15, Last sequence update)   |      |          |  |  |  |
| DT       | 01-OCT-2003 (TREMBrel. 25, Last annotation update)   |      |          |  |  |  |
| DE       | Genome polyprotein.  |      |          |  |  |  |
| OS       | Repartitis C virus.  |      |          |  |  |  |
| OC       | Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  |      |          |  |  |  |
| OX       | Repavirus.   |      |          |  |  |  |
| RN       | [1] - NCBI_TAXID=1103;   |      |          |  |  |  |
| RP       | SEQUENCE FROM N.A.   |      |          |  |  |  |
| RC       | STRAIN=M03;  |      |          |  |  |  |
| RA       | Nagayama K., Kuroski M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;  |      |          |  |  |  |
| RT       | "Characteristics of hepatitis C viral genome associated with disease progression,"   |      |          |  |  |  |
| RT       | Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  |      |          |  |  |  |
| CC       | -- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLOCCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLICOPROTEIN E. THE NUCLOECAPSID IS A COMPLEX OF PROTEIN C AND M RNA (BY SIMILARITY). |      |          |  |  |  |
| EMBL     | EMBL: AF02774; AA165964.1; -   |      |          |  |  |  |
| DR       | PIR; AG1196; AG1196.   |      |          |  |  |  |
| DR       | PIR; PQ0246; PQ0329.   |      |          |  |  |  |
| DR       | PS0319; PS0329.  |      |          |  |  |  |
| DR       | HSSP; P27958; IHBI.  |      |          |  |  |  |
| DR       | MEROPS; S28.001; -.  |      |          |  |  |  |
| DR       | MEROPS; U39.001; -.  |      |          |  |  |  |
| DR       | GO; GO:0016021; C:integral to membrane; IEA.   |      |          |  |  |  |
| DR       | GO; GO:0019028; C:viral capsid; IEA.   |      |          |  |  |  |
| DR       | GO; GO:0019031; C:viral envelope; IEA.   |      |          |  |  |  |
| DR       | GO; GO:0005524; F:AMP binding; IEA.  |      |          |  |  |  |
| DR       | GO; GO:0008026; F:ATP dependent helicase activity; IEA.  |      |          |  |  |  |
| DR       | GO; GO:0005489; F:electron transporter activity; IEA.  |      |          |  |  |  |
| DR       | GO; GO:0003723; F:RNA binding; IEA.  |      |          |  |  |  |
| DR       | GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.   |      |          |  |  |  |
| DR       | GO; GO:0008236; F:serine-type peptidase activity; IEA.   |      |          |  |  |  |



DR Pfam; PF00998; Viral\_RdRP; 1.  
 DR ProDom; PD18602; HCV\_NSI; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 DR PROSITE; PS00190; CYTOCHROME\_C; 1.  
 KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;  
 KW Hydrolase; Nonstructural protein; Polyprotein;  
 KW RNA-directed RNA polymerase; Transferase; Transmembrane;  
 SQ SEQUENCE 3010 AA; 327042 NM; 307DCB979884C95 CRC64;

Qy 2 LTLSPPYYKVLARLWILQYLITRVEAH!QW!IPPLNVRGGDAIILTCAVHPELID 61  
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
 828 LLLTSPYYKVFLAKLWILQYLITRVEAH!QW!IPPLNVRGGDAIILTCAVHPELID 887  
 DR GO; GO:0019031; C:viral capsid; IEA.  
 DR GO; GO:0019031; C:viral envelope; IEA.  
 DR GO; GO:000524; F:ATP binding; IEA.  
 DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.  
 DR GO; GO:0005889; F:Electron transporter activity; IEA.  
 DR GO; GO:000323; F:RNA binding; IEA.  
 DR GO; GO:003568; F:RNA directed RNA polymerase activity; IEA.  
 DR GO; GO:008236; F:serine-type peptidase activity; IEA.  
 DR GO; GO:0005198; F:Structural molecule activity; IEA.  
 DR GO; GO:0016740; F:Transferase activity; IEA.  
 DR GO; GO:006118; P:Electron transport; IEA.  
 DR GO; GO:0006508; P:Proteolysis and peptidolysis; IEA.  
 DR GO; GO:0006550; P:transcription; IEA.  
 DR GO; GO:0019079; P:viral genome replication; IEA.  
 DR GO; GO:0019087; P:viral transformation; IEA.  
 DR InterPro; IPR00903; Cys\_Ser\_trypsin.  
 DR InterPro; IPR00335; Cyt\_C\_heme\_BS.  
 DR InterPro; IPR00140; DEAD.  
 DR InterPro; IPR00252; HCV capsid.  
 DR InterPro; IPR00251; HCV core.  
 DR InterPro; IPR00259; HCV env.  
 DR InterPro; IPR002511; HCV NS1.  
 DR InterPro; IPR002518; HCV NS2.  
 DR InterPro; IPR000715; HCV\_NSA.  
 DR InterPro; IPR00140; HCV\_NSB.  
 DR InterPro; IPR002868; HCV\_NSS.  
 DR InterPro; IPR002511; HCV\_RdP.  
 DR InterPro; IPR004109; Peptidase\_C9.  
 DR InterPro; IPR00705; RNA\_pol\_DS\_Ps.  
 DR InterPro; IPR00704; RNA\_pol\_PSVir.  
 DR Pfam; PF01543; HCV capsid; 1.  
 DR Pfam; PF01542; HCV core; 1.  
 DR Pfam; PF01539; HCV\_RdP.  
 DR Pfam; PF01560; HCV\_NSI; 1.  
 DR Pfam; PF01538; HCV\_NS2; 1.  
 DR Pfam; PF02907; HCV\_NSA; 1.  
 DR Pfam; PF01006; HCV\_NSB; 1.  
 DR Pfam; PF01001; HCV\_NS4b; 1.  
 DR Pfam; PF01506; HCV\_NS5a; 1.  
 DR Pfam; PF00998; viral\_RdP; 1.  
 DR ProDom; PD18602; HCV\_NSI; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 DR PROSITE; PS00190; CYTOCHROME\_C; 1.  
 DR Polypeptide; Envelope protein; Glycoprotein; Nonstructural protein;  
 KW Protein; RNA-directed RNA polymerase; Transferase; Transmembrane.  
 SQ SEQUENCE 3010 AA; 327365 NM; D865F7317FFA06 CRC64;

Qy 09J3H7 PRELIMINARY; PRT; 3010 AA.  
 AC 09J3H7  
 DT 01-OCT-2000 (Tremblet, 15, Created)  
 DT 01-OCT-2000 (Tremblet, 15, Last sequence update)  
 DT 01-OCT-2003 (Tremblet, 25, Last annotation update)  
 DB Genome\_polyprotein.

OS Hepatitis C Virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OX NCBI\_TAXID=1103;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN=M15;  
 RA Nagayama K., Kuroasaki M., Enomoto N., Miyazaka Y., Marumo F., Sato C.,  
 RT "characteristics of hepatitis C viral genome associated with disease  
 progression."  
 RT Submitted (Nov-1999) to the EMBL/GenBank/DBJ databases.  
 CC - SUBMIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:  
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  
 PROTEIN C AND MRNA (BY SIMILARITY).  
 DR EMBL; AF207756; AF65946.1; -;  
 DR PIR; A6119; A61196.  
 DR PIR; PQ0246; PQ0246.  
 DR PIR; PQ0804.  
 DR PIR; PS0329; PS0329.  
 DR HSP; P26663; IJXP.  
 DR GO; GO:0016021; C:integral to membrane; IEA.

DR GO; GO:0019028; C:viral capsid; IEA.  
 DR GO; GO:0019031; C:viral envelope; IEA.  
 DR GO; GO:000524; F:ATP binding; IEA.  
 DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.  
 DR GO; GO:0005889; F:Electron transporter activity; IEA.  
 DR GO; GO:000323; F:RNA binding; IEA.  
 DR GO; GO:003568; F:RNA directed RNA polymerase activity; IEA.  
 DR GO; GO:008236; F:serine-type peptidase activity; IEA.  
 DR GO; GO:0005198; F:Structural molecule activity; IEA.  
 DR GO; GO:0016740; F:Transferase activity; IEA.  
 DR GO; GO:006118; P:Electron transport; IEA.  
 DR GO; GO:0006508; P:Proteolysis and peptidolysis; IEA.  
 DR GO; GO:0006550; P:transcription; IEA.  
 DR GO; GO:0019079; P:viral genome replication; IEA.  
 DR GO; GO:0019087; P:viral transformation; IEA.  
 DR InterPro; IPR00903; Cys\_Ser\_trypsin.  
 DR InterPro; IPR00335; Cyt\_C\_heme\_BS.  
 DR InterPro; IPR00140; DEAD.  
 DR InterPro; IPR00252; HCV capsid.  
 DR InterPro; IPR00251; HCV core.  
 DR InterPro; IPR00259; HCV env.  
 DR InterPro; IPR002511; HCV NS1.  
 DR InterPro; IPR002518; HCV NS2.  
 DR InterPro; IPR000715; HCV\_NSA.  
 DR InterPro; IPR00140; HCV\_NSB.  
 DR InterPro; IPR002868; HCV\_NSS.  
 DR InterPro; IPR002511; HCV\_RdP.  
 DR InterPro; IPR004109; Peptidase\_C9.  
 DR InterPro; IPR00705; RNA\_pol\_DS\_Ps.  
 DR InterPro; IPR00704; RNA\_pol\_PSVir.  
 DR Pfam; PF01543; HCV capsid; 1.  
 DR Pfam; PF01542; HCV core; 1.  
 DR Pfam; PF01539; HCV\_RdP.  
 DR Pfam; PF01560; HCV\_NSI; 1.  
 DR Pfam; PF01538; HCV\_NS2; 1.  
 DR Pfam; PF02907; HCV\_NSA; 1.  
 DR Pfam; PF01006; HCV\_NSB; 1.  
 DR Pfam; PF01001; HCV\_NS4b; 1.  
 DR Pfam; PF01506; HCV\_NS5a; 1.  
 DR Pfam; PF00998; viral\_RdP; 1.  
 DR ProDom; PD18602; HCV\_NSI; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 DR PROSITE; PS00190; CYTOCHROME\_C; 1.  
 DR Polypeptide; Envelope protein; Glycoprotein; Nonstructural protein;  
 KW Protein; RNA-directed RNA polymerase; Transferase; Transmembrane.  
 SQ SEQUENCE 3010 AA; 327365 NM; D865F7317FFA06 CRC64;

Qy 1 AUTLSPYYKVLARLWILQYLITRVEAH!QW!IPPLNVRGGDAIILTCAVHPELID 60  
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
 827 ALLTSPYYKVFLAKLWILQYLITRVEAH!QW!IPPLNVRGGDAIILTCAVHPELID 886  
 DR GO; GO:0019031; C:viral capsid; IEA.  
 DR GO; GO:0019031; C:viral envelope; IEA.  
 DR GO; GO:000524; F:ATP binding; IEA.  
 DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.  
 DR GO; GO:0005889; F:Electron transporter activity; IEA.  
 DR GO; GO:000323; F:RNA binding; IEA.  
 DR GO; GO:003568; F:RNA directed RNA polymerase activity; IEA.  
 DR GO; GO:008236; F:serine-type peptidase activity; IEA.  
 DR GO; GO:0005198; F:Structural molecule activity; IEA.  
 DR GO; GO:0016740; F:Transferase activity; IEA.  
 DR GO; GO:006118; P:Electron transport; IEA.  
 DR GO; GO:0006508; P:Proteolysis and peptidolysis; IEA.  
 DR GO; GO:0006550; P:transcription; IEA.  
 DR GO; GO:0019079; P:viral genome replication; IEA.  
 DR GO; GO:0019087; P:viral transformation; IEA.  
 DR InterPro; IPR00903; Cys\_Ser\_trypsin.  
 DR InterPro; IPR00335; Cyt\_C\_heme\_BS.  
 DR InterPro; IPR00140; DEAD.  
 DR InterPro; IPR00252; HCV capsid.  
 DR InterPro; IPR00251; HCV core.  
 DR InterPro; IPR00259; HCV env.  
 DR InterPro; IPR002511; HCV NS1.  
 DR InterPro; IPR002518; HCV NS2.  
 DR InterPro; IPR000715; HCV\_NSA.  
 DR InterPro; IPR00140; HCV\_NSB.  
 DR InterPro; IPR002868; HCV\_NSS.  
 DR InterPro; IPR002511; HCV\_RdP.  
 DR InterPro; IPR004109; Peptidase\_C9.  
 DR InterPro; IPR00705; RNA\_pol\_DS\_Ps.  
 DR InterPro; IPR00704; RNA\_pol\_PSVir.  
 DR Pfam; PF01543; HCV capsid; 1.  
 DR Pfam; PF01542; HCV core; 1.  
 DR Pfam; PF01539; HCV\_RdP.  
 DR Pfam; PF01560; HCV\_NSI; 1.  
 DR Pfam; PF01538; HCV\_NS2; 1.  
 DR Pfam; PF02907; HCV\_NSA; 1.  
 DR Pfam; PF01006; HCV\_NSB; 1.  
 DR Pfam; PF01001; HCV\_NS4b; 1.  
 DR Pfam; PF01506; HCV\_NS5a; 1.  
 DR Pfam; PF00998; viral\_RdP; 1.  
 DR ProDom; PD18602; HCV\_NSI; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 DR PROSITE; PS00190; CYTOCHROME\_C; 1.  
 DR Polypeptide; Envelope protein; Glycoprotein; Nonstructural protein;  
 KW Protein; RNA-directed RNA polymerase; Transferase; Transmembrane.  
 SQ SEQUENCE 3010 AA; 327365 NM; D865F7317FFA06 CRC64;

Qy 121 TYVYDHLTPLQDWAHGLRDLAVAVEPVFSMDMEVKIITWGAQDAAAGGCGCITSLGRDNQFEGEVQSTAT 180  
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
 827 ALLTSPYYKVFLAKLWILQYLITRVEAH!QW!IPPLNVRGGDAIILTCAVHPELID 886  
 DR GO; GO:0019031; C:viral capsid; IEA.  
 DR GO; GO:0019031; C:viral envelope; IEA.  
 DR GO; GO:000524; F:ATP binding; IEA.  
 DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.  
 DR GO; GO:0005889; F:Electron transporter activity; IEA.  
 DR GO; GO:000323; F:RNA binding; IEA.  
 DR GO; GO:003568; F:RNA directed RNA polymerase activity; IEA.  
 DR GO; GO:008236; F:serine-type peptidase activity; IEA.  
 DR GO; GO:0005198; F:Structural molecule activity; IEA.  
 DR GO; GO:0016740; F:Transferase activity; IEA.  
 DR GO; GO:006118; P:Electron transport; IEA.  
 DR GO; GO:0006508; P:Proteolysis and peptidolysis; IEA.  
 DR GO; GO:0006550; P:transcription; IEA.  
 DR GO; GO:0019079; P:viral genome replication; IEA.  
 DR GO; GO:0019087; P:viral transformation; IEA.  
 DR InterPro; IPR00903; Cys\_Ser\_trypsin.  
 DR InterPro; IPR00335; Cyt\_C\_heme\_BS.  
 DR InterPro; IPR00140; DEAD.  
 DR InterPro; IPR00252; HCV capsid.  
 DR InterPro; IPR00251; HCV core.  
 DR InterPro; IPR00259; HCV env.  
 DR InterPro; IPR002511; HCV NS1.  
 DR InterPro; IPR002518; HCV NS2.  
 DR InterPro; IPR000715; HCV\_NSA.  
 DR InterPro; IPR00140; HCV\_NSB.  
 DR InterPro; IPR002868; HCV\_NSS.  
 DR InterPro; IPR002511; HCV\_RdP.  
 DR InterPro; IPR004109; Peptidase\_C9.  
 DR InterPro; IPR00705; RNA\_pol\_DS\_Ps.  
 DR InterPro; IPR00704; RNA\_pol\_PSVir.  
 DR Pfam; PF01543; HCV capsid; 1.  
 DR Pfam; PF01542; HCV core; 1.  
 DR Pfam; PF01539; HCV\_RdP.  
 DR Pfam; PF01560; HCV\_NSI; 1.  
 DR Pfam; PF01538; HCV\_NS2; 1.  
 DR Pfam; PF02907; HCV\_NSA; 1.  
 DR Pfam; PF01006; HCV\_NSB; 1.  
 DR Pfam; PF01001; HCV\_NS4b; 1.  
 DR Pfam; PF01506; HCV\_NS5a; 1.  
 DR Pfam; PF00998; viral\_RdP; 1.  
 DR ProDom; PD18602; HCV\_NSI; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 DR PROSITE; PS00190; CYTOCHROME\_C; 1.  
 DR Polypeptide; Envelope protein; Glycoprotein; Nonstructural protein;  
 KW Protein; RNA-directed RNA polymerase; Transferase; Transmembrane.  
 SQ SEQUENCE 3010 AA; 327365 NM; D865F7317FFA06 CRC64;

Qy 241 OSFLATCNGVWCTVPHGASSTKLQGPKGTQITOMTNVDODLVGQAPPGRSMPCTCG 300  
 |||||:|||||:|||||:  
 827 ALLTSPYYKVFLAKLWILQYLITRVEAH!QW!IPPLNVRGGDAIILTCAVHPELID 886  
 DR GO; GO:0019031; C:viral capsid; IEA.  
 DR GO; GO:0019031; C:viral envelope; IEA.  
 DR GO; GO:000524; F:ATP binding; IEA.  
 DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.  
 DR GO; GO:0005889; F:Electron transporter activity; IEA.  
 DR GO; GO:000323; F:RNA binding; IEA.  
 DR GO; GO:003568; F:RNA directed RNA polymerase activity; IEA.  
 DR GO; GO:008236; F:serine-type peptidase activity; IEA.  
 DR GO; GO:0005198; F:Structural molecule activity; IEA.  
 DR GO; GO:0016740; F:Transferase activity; IEA.  
 DR GO; GO:006118; P:Electron transport; IEA.  
 DR GO; GO:0006508; P:Proteolysis and peptidolysis; IEA.  
 DR GO; GO:0006550; P:transcription; IEA.  
 DR GO; GO:0019079; P:viral genome replication; IEA.  
 DR GO; GO:0019087; P:viral transformation; IEA.  
 DR InterPro; IPR00903; Cys\_Ser\_trypsin.  
 DR InterPro; IPR00335; Cyt\_C\_heme\_BS.  
 DR InterPro; IPR00140; DEAD.  
 DR InterPro; IPR00252; HCV capsid.  
 DR InterPro; IPR00251; HCV core.  
 DR InterPro; IPR00259; HCV env.  
 DR InterPro; IPR002511; HCV NS1.  
 DR InterPro; IPR002518; HCV NS2.  
 DR InterPro; IPR000715; HCV\_NSA.  
 DR InterPro; IPR00140; HCV\_NSB.  
 DR InterPro; IPR002868; HCV\_NSS.  
 DR InterPro; IPR002511; HCV\_RdP.  
 DR InterPro; IPR004109; Peptidase\_C9.  
 DR InterPro; IPR00705; RNA\_pol\_DS\_Ps.  
 DR InterPro; IPR00704; RNA\_pol\_PSVir.  
 DR Pfam; PF01543; HCV capsid; 1.  
 DR Pfam; PF01542; HCV core; 1.  
 DR Pfam; PF01539; HCV\_RdP.  
 DR Pfam; PF01560; HCV\_NSI; 1.  
 DR Pfam; PF01538; HCV\_NS2; 1.  
 DR Pfam; PF02907; HCV\_NSA; 1.  
 DR Pfam; PF01006; HCV\_NSB; 1.  
 DR Pfam; PF01001; HCV\_NS4b; 1.  
 DR Pfam; PF01506; HCV\_NS5a; 1.  
 DR Pfam; PF00998; viral\_RdP; 1.  
 DR ProDom; PD18602; HCV\_NSI; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 DR PROSITE; PS00190; CYTOCHROME\_C; 1.  
 DR Polypeptide; Envelope protein; Glycoprotein; Nonstructural protein;  
 KW Protein; RNA-directed RNA polymerase; Transferase; Transmembrane.  
 SQ SEQUENCE 3010 AA; 327365 NM; D865F7317FFA06 CRC64;

Qy 1067 OSFLATCNGVWCTVPHGASSTKLQGPKGTQITOMTNVDODLVGQAPPGRSMPCTCG 1126  
 |||||:|||||:  
 827 ALLTSPYYKVFLAKLWILQYLITRVEAH!QW!IPPLNVRGGDAIILTCAVHPELID 886  
 DR GO; GO:0019031; C:viral capsid; IEA.  
 DR GO; GO:0019031; C:viral envelope; IEA.  
 DR GO; GO:000524; F:ATP binding; IEA.  
 DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.  
 DR GO; GO:0005889; F:Electron transporter activity; IEA.  
 DR GO; GO:000323; F:RNA binding; IEA.  
 DR GO; GO:003568; F:RNA directed RNA polymerase activity; IEA.  
 DR GO; GO:008236; F:serine-type peptidase activity; IEA.  
 DR GO; GO:0005198; F:Structural molecule activity; IEA.  
 DR GO; GO:0016740; F:Transferase activity; IEA.  
 DR GO; GO:006118; P:Electron transport; IEA.  
 DR GO; GO:0006508; P:Proteolysis and peptidolysis; IEA.  
 DR GO; GO:0006550; P:transcription; IEA.  
 DR GO; GO:0019079; P:viral genome replication; IEA.  
 DR GO; GO:0019087; P:viral transformation; IEA.  
 DR InterPro; IPR00903; Cys\_Ser\_trypsin.  
 DR InterPro; IPR00335; Cyt\_C\_heme\_BS.  
 DR InterPro; IPR00140; DEAD.  
 DR InterPro; IPR00252; HCV capsid.  
 DR InterPro; IPR00251; HCV core.  
 DR InterPro; IPR00259; HCV env.  
 DR InterPro; IPR002511; HCV NS1.  
 DR InterPro; IPR002518; HCV NS2.  
 DR InterPro; IPR000715; HCV\_NSA.  
 DR InterPro; IPR00140; HCV\_NSB.  
 DR InterPro; IPR002868; HCV\_NSS.  
 DR InterPro; IPR002511; HCV\_RdP.  
 DR InterPro; IPR004109; Peptidase\_C9.  
 DR InterPro; IPR00705; RNA\_pol\_DS\_Ps.  
 DR InterPro; IPR00704; RNA\_pol\_PSVir.  
 DR Pfam; PF01543; HCV capsid; 1.  
 DR Pfam; PF01542; HCV core; 1.  
 DR Pfam; PF01539; HCV\_RdP.  
 DR Pfam; PF01560; HCV\_NSI; 1.  
 DR Pfam; PF01538; HCV\_NS2; 1.  
 DR Pfam; PF02907; HCV\_NSA; 1.  
 DR Pfam; PF01006; HCV\_NSB; 1.  
 DR Pfam; PF01001; HCV\_NS4b; 1.  
 DR Pfam; PF01506; HCV\_NS5a; 1.  
 DR Pfam; PF00998; viral\_RdP; 1.  
 DR ProDom; PD18602; HCV\_NSI; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 DR PROSITE; PS00190; CYTOCHROME\_C; 1.  
 DR Polypeptide; Envelope protein; Glycoprotein; Nonstructural protein;  
 KW Protein; RNA-directed RNA polymerase; Transferase; Transmembrane.  
 SQ SEQUENCE 3010 AA; 327365 NM; D865F7317FFA06 CRC64;

Qy 301 SSDLYVTRHADIVPYRGRPSRGSLISPRPVSYLKGSQGCPPLCPGSHAVGIFRAVCT 360  
 |||||:|||||:  
 827 ALLTSPYYKVFLAKLWILQYLITRVEAH!QW!IPPLNVRGGDAIILTCAVHPELID 886

|  |  |   |      |  |
|--|--|---|------|--|
| Db   | 1127                                   | SSDLVLTAVHADVIPVRRGDSRGSLLSPRVSYLKGGGGPLCPSGHAVGIFRAVCT | 1186 | DR Pfam; PF01560; HCV_NS1; 1.  |
| OY   | 361                                    | RGVAKAVDIPVESMETWR                                      | 380  | DR Pfam; PF01538; HCV_NS2; 1.  |
| Db   | 1187                                   | RGVAKAVDIPVESMETWR                                      | 1206 | DR Pfam; PF02907; HCV_NS3; 1.  |
|  |  |   |      | DR Pfam; PF01006; HCV_NS4; 1.  |
|  |  |   |      | DR Pfam; PF01001; HCV_NS4b; 1.   |
| RESULT 4   |  |   |      | DR Pfam; PF01506; HCV_NS5a; 1.   |
| Q9DTD6   |  | PRELIMINARY;  | PRT; | DR Pfam; PF00271; helicase_C; 1.   |
| ID Q9DTD6  |  |   |      | DR Pfam; PF00998; Viral_RdRP; 1.   |
| AC Q9DTD6  |  |   |      | DR PRODom; PDB186062; HCV_NS1; 1.  |
| DT 01-MAR-2001   | (TREMBrel. 16, Created)                |   |      | DR SMART; SW00487; DEXD_C; 1.  |
| DT 01-MAR-2001   | (TREMBrel. 16, Last sequence update)   |   |      | DR PROSITE; PS0190; CTOCHROMEC; 1.   |
| DT 01-OCT-2003   | (TREMBrel. 25, Last annotation update) |   |      | DR ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;  |
| DE Genome_Polyprotein.   |  |   |      | DR Hydrolase; Nonstructural protein; Polyprotein;  |
| OS Hepatitis_C_virus.  |  |   |      | DR RNA-directed RNA_polymerase; Transferase; Transmembrane.  |
| OC Viruses; ssRNA_positive-strand viruses, no DNA stage; Flaviviridae;   |  |   |      | DR Sequence FROM N.A.  |
| OC Hepatitis_Virus.  |  |   |      | DR STRAIN_HCVt21;  |
| OX NCBI_TAXID=1103;  |  |   |      | DR Takaishi K., Iwata K., Matsumoto M., Matsumoto H., Nakao K.,  |
| RN [1]   |  |   |      | DR RA Hatahara T., Ohta Y., Kanai K., Maruo H., Baba K., Hijikata M.,  |
| RP   |  |   |      | DR Mishiro S.,   |
| RC   |  |   |      | DR "Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients with hepatocellular carcinoma: the 'progression score' revisited"; |
| RA   |  |   |      | DR Submitted (SBR-2000) to the EMBL/GenBank/DDBJ databases.  |
| RT -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND M/RNA (BY SIMILARITY). |  |   |      | DR CC EMBL; AB049101; BAB1814_1; -   |
| RL PTR; A61196; A61195.  |  |   |      | DR DR P00246; P00246;  |
| CC PIR; P00329; PS029.   |  |   |      | DR DR P00329; PS029.   |
| HASP; P26663; 1xP.   |  |   |      | DR DR HASP; P26663; 1xP.   |
| GO; GO_0016021; C:integral_to_membrane; IEA.   |  |   |      | DR GO; GO_0016021; C:viral_capsid; IEA.  |
| DR GO; GO_0019028; C:viral_capsid; IEA.  |  |   |      | DR GO; GO_0019028; C:viral_capsid; IEA.  |
| DR GO; GO_001931; C:viral_envelope; IEA.   |  |   |      | DR GO; GO_001931; C:viral_envelope; IEA.   |
| DR GO; GO_000552; F:ATP_binding; IEA.  |  |   |      | DR GO; GO_000552; F:ATP_binding; IEA.  |
| DR GO; GO_0008026; F:ATP_dependent_helicase_activity; IEA.   |  |   |      | DR GO; GO_0008026; F:ATP_dependent_helicase_activity; IEA.   |
| DR GO; GO_0010489; F:electron_transporter_activity; IEA.   |  |   |      | DR GO; GO_0010489; F:electron_transporter_activity; IEA.   |
| DR GO; GO_0011678; F:hydrolase_activity; IEA.  |  |   |      | DR GO; GO_0011678; F:hydrolase_activity; IEA.  |
| DR GO; GO_001323; F:RNA_binding; IEA.  |  |   |      | DR GO; GO_001323; F:RNA_binding; IEA.  |
| DR GO; GO_001618; F:RNA-directed RNA_polymerase_activity; IEA.   |  |   |      | DR GO; GO_001618; F:RNA-directed RNA_polymerase_activity; IEA.   |
| DR GO; GO_001936; F:serine-type_peptidase_activity; IEA.   |  |   |      | DR GO; GO_001936; F:serine-type_peptidase_activity; IEA.   |
| DR GO; GO_001938; F:structural_molecule_activity; IEA.   |  |   |      | DR GO; GO_001938; F:structural_molecule_activity; IEA.   |
| DR GO; GO_0016740; F:tRNA_transport; IEA.  |  |   |      | DR GO; GO_0016740; F:tRNA_transport; IEA.  |
| DR GO; GO_001618; P:electron_transport; IEA.   |  |   |      | DR GO; GO_001618; P:electron_transport; IEA.   |
| DR GO; GO_0005508; P:proteolysis_and_peptidolysis; IEA.  |  |   |      | DR GO; GO_0005508; P:proteolysis_and_peptidolysis; IEA.  |
| DR GO; GO_0003350; P:transcription; IEA.   |  |   |      | DR GO; GO_0003350; P:transcription; IEA.   |
| DR GO; GO_0019079; P:viral_genome_rePLICATION; IEA.  |  |   |      | DR GO; GO_0019079; P:viral_genome_rePLICATION; IEA.  |
| DR GO; GO_0019087; P:viral_transformation; IEA.  |  |   |      | DR GO; GO_0019087; P:viral_transformation; IEA.  |
| DR InterPro; IPR003003; Cys_Ser_trypsin.   |  |   |      | DR InterPro; IPR003003; Cys_Ser_trypsin.   |
| DR InterPro; IPR00345; CytC_heme_BS.   |  |   |      | DR InterPro; IPR00345; CytC_heme_BS.   |
| DR InterPro; IPR001410; DEAD.  |  |   |      | DR InterPro; IPR001410; DEAD.  |
| DR InterPro; IPR003521; HCV_ns1.   |  |   |      | DR InterPro; IPR003521; HCV_ns1.   |
| DR InterPro; IPR002519; HCV_env.   |  |   |      | DR InterPro; IPR002519; HCV_env.   |
| DR InterPro; IPR02531; HCV_N51.  |  |   |      | DR InterPro; IPR02531; HCV_N51.  |
| DR InterPro; IPR00518; HCV_N52.  |  |   |      | DR InterPro; IPR00518; HCV_N52.  |
| DR InterPro; IPR000745; HCV_ns4a.  |  |   |      | DR InterPro; IPR000745; HCV_ns4a.  |
| DR InterPro; IPR00190; HCV_ns4b.   |  |   |      | DR InterPro; IPR00190; HCV_ns4b.   |
| DR InterPro; IPR002868; HCV_N5a.   |  |   |      | DR InterPro; IPR002868; HCV_N5a.   |
| DR InterPro; IPR002166; HCV_RdRP.  |  |   |      | DR InterPro; IPR002166; HCV_RdRP.  |
| DR InterPro; IPR001650; Helicase_C.  |  |   |      | DR InterPro; IPR001650; Helicase_C.  |
| DR InterPro; IPR004109; Pepidas_C29.   |  |   |      | DR InterPro; IPR004109; Pepidas_C29.   |
| DR InterPro; IPR007095; RNA_poli_Ds_Ps.  |  |   |      | DR InterPro; IPR007095; RNA_poli_Ds_Ps.  |
| DR Pfam; PF01543; HCV_capsid; 1.   |  |   |      | DR Pfam; PF01543; HCV_capsid; 1.   |
| DR Pfam; PF01542; HCV_core; 1.   |  |   |      | DR Pfam; PF01542; HCV_core; 1.   |
| DR Pfam; PF01539; HCV_env; 1.  |  |   |      | DR Pfam; PF01539; HCV_env; 1.  |

|                       |  |    |   |
|-----------------------|--|----|---|
| CC                    | PROTEIN C AND mRNA (BY SIMILARITY).  | QY | 181 GREB1LGPADNEFGQWRLAPITAYSQTRGLGCLTSLGDRDNQKVGEQVSTAT 240        |
| DR                    | EMBL; AB41909; BA11806.1; -.   | Db | 1007 GKEILGPADSLRGQGRLLAPITAYSQTRGLGCLTSLGDRDNQKVGEQVSTAT 1066      |
| DR                    | PIR; A61196; A61196.   | QY | 241 QSLFLATCNGVGVTCVHAGSKTLAGKGPITQMYTNUDQLVGMQAPRGAARSMPCTCG 300   |
| DR                    | PIR; PS0246; PS0246.   | Db | 1067 QSLFLATCNGVGVTCVHAGSKTLAGKGPITQMYTNUDQLVGMQAPRGAARSMPCTCG 1126 |
| DR                    | PIR; PS0249; PS0329.   | QY | 301 SSDIYLVYRHADYIPVRRGDSFGSLSPPRPSYVLKSSGGPLCPSGHAVGIFRAVCT 360    |
| DR                    | HSSP; P26663; 1JXP.  | Db | 1127 SSDIYLVYRHADYIPVRRGDSFGSLSPPRPSYVLKSSGGPLCPSGHAVGIFRAVCT 1186  |
| DR                    | GO; GO:0016021; C:integral to membrane; IEA.   | QY | 361 RGYAKAVDFIPVSMETMR 380  |
| DR                    | GO; GO:0019028; C:viral capsid; IEA.   | Db | 1187 RGVAKAVDFPVVESMETMR 1206                                       |
| DR                    | GO; GO:0019031; C:viral envelope; IEA.   |    |   |
| DR                    | GO; GO:0005198; F:structural molecule activity; IEA.   |    |   |
| DR                    | GO; GO:016740; F:transferase activity; IEA.  |    |   |
| DR                    | GO; GO:0006118; F:electron transport; IEA.   |    |   |
| DR                    | GO; GO:0006508; P:proteolysis and peptidolysis; IEA..  |    |   |
| DR                    | GO; GO:0006350; P:transcription; IEA..   |    |   |
| DR                    | GO; GO:0019079; P:viral genome replication; IEA.   |    |   |
| DR                    | GO; GO:0019087; P:viral transformation; IEA.   |    |   |
| DR                    | InterPro; IPR009003; Cys_Ser_trypsin.  |    |   |
| DR                    | InterPro; IPR00345; CYT_C_heme_BS.   |    |   |
| DR                    | InterPro; IPR001410; DEAD.   |    |   |
| DR                    | InterPro; IPR002522; HCV capsid.   |    |   |
| DR                    | InterPro; IPR02521; HCV_core.  |    |   |
| DR                    | InterPro; IPR02519; HCV_env.   |    |   |
| DR                    | InterPro; IPR02518; HCV_NS1.   |    |   |
| DR                    | InterPro; IPR000745; HCV_NS4a.   |    |   |
| DR                    | InterPro; IPR002868; HCV_NS5a.   |    |   |
| DR                    | InterPro; IPR001490; HCV_NS4b.   |    |   |
| DR                    | InterPro; IPR001650; HCV_RdRp.   |    |   |
| DR                    | InterPro; IPR004109; Peptidase_C29.  |    |   |
| DR                    | InterPro; IPR007051; RNA_Pol_DS_PS.  |    |   |
| DR                    | InterPro; IPR007094; RNA_Pol_PSVir.  |    |   |
| DR                    | Pfam; PF01543; HCV_capsid; 1.  |    |   |
| DR                    | Pfam; PF01442; HCV_core; 1.  |    |   |
| DR                    | Pfam; PF01339; HCV_env; 1.   |    |   |
| DR                    | Pfam; PF0160; HCV_NS1; 1.  |    |   |
| DR                    | Pfam; PF0138; HCV_NS2; 1.  |    |   |
| DR                    | Pfam; PF02907; HCV_NS3; 1.   |    |   |
| DR                    | Pfam; PF01005; HCV_NS4a; 1.  |    |   |
| DR                    | Pfam; PF01001; HCV_NS4b; 1.  |    |   |
| DR                    | Pfam; PF01566; HCV_NSSa; 1.  |    |   |
| DR                    | Pfam; PF00271; Helicase_C; 1.  |    |   |
| DR                    | Pfam; PP00988; Viral_RdRp; 1.  |    |   |
| DR                    | ProDom; PDB186062; HCV_NS1; 1.   |    |   |
| DR                    | SMART; SM00487; DEXDC; 1.  |    |   |
| DR                    | SMART; SM00190; CYTOCHROME_C; 1.   |    |   |
| KW                    | Cat protein; Envelope protein; Glycoprotein; Nonsstructural protein; Polyprotein; RNA-directed RNA_polymerase; Transferase; Transmembrane. |    |   |
| SQ                    | SEQUENCE 3010 AA; 327324 MW; 3DB6CF249BD1151C CRC64;   |    |   |
| Query Match           | 96.0%; Score 1908; DB 12; Length 3010;   |    |   |
| Best Local Similarity | 94.5%; Prd. No. 6_9e-152;  |    |   |
| Matches               | 359; Conservative 10; Mismatches 11; Indels 0; Gaps 0;   |    |   |
| QY                    | 1 ALITLSPYKVLLARLWILQYLITRVEAHQWIPPLVNRGRDAILITCAVHPFLF 60   |    |   |
| Db                    | 827 ALITLSPYKVFLARLWILQYLITRVEAHQWIPPLVNRGRDAILITCAVHPFLF 886  |    |   |
| QY                    | 61 DITKULLAIFGPMVLYQALITKTVYFVRAGQLRACMVVKRKAAGGHYQMANVKAALIG 120  |    |   |
| Db                    | 887 DITKULLAIFGPMVLYQALITKTVYFVRAGQLRACMVVKRKAAGGHYQMANVKAALIG 946   |    |   |
| QY                    | 121 YYVVDLHTLQDWHAAGLDRALAVAVEPVFSDEMEVKITWTGAGTAACDDITSLPVSARR 180  |    |   |
| Db                    | 947 YYVVDLHTLQDWHAASRUDLAVAVEPVFSDEMEVKITWTGAGTAACDDITSLPVSARR 1006  |    |   |
| DR                    | Pfam; PF01539; HCV_env; 1.   |    |   |
| DR                    | Pfam; PF01560; HCV_NS1; 1.   |    |   |



**RESULT 8**

ID Q68788 PRELIMINARY; PRT; 3010 AA.

AC Q68788;

DT 01-NOV-1995 (TREMBrel. 01, Created)

DT 01-NOV-1995 (TREMBrel. 01, Last sequence update)

DT 01-OCT-2003 (TREMBrel. 25, Last annotation update)

DE HCV polyprotein (Genome polyprotein).

OS Hepatitis C virus.

VIRUSES; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepatitis C virus.

OC HCV.

OX NCBI\_TAXID:1103;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=96362156; PubMed=8720135;

RA Seki M., Henda Y.,

RT "Phosphorochioate antisense oligodeoxynucleotides capable of inhibiting Hepatitis C virus gene expression: In vitro translation assay."

J. Biochem. 118:1199-1204 (1995). THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND mRNA (BY SIMILARITY).

CC EMBL; D45172; BA08120.1; -.

DR PIR; A61196; A61196.

DR PIR; P00246; P00246.

DR PIR; PS0329; PS0329.

DR RSSP; P26663; IUP.

DR GO; GO\_0016021; C:integral membrane protein; IEA.

DR GO; GO\_0019028; C:viral capsid; IEA.

DR GO; GO\_0019031; C:viral envelope; IEA.

DR GO; GO\_005524; F:ATP binding; IEA.

DR GO; GO\_0008026; F:ATP dependent helicase activity; IEA.

DR GO; GO\_0003723; F:RNA binding; IEA.

DR GO; GO\_000968; F:RNA-directed RNA polymerase activity; IEA.

DR GO; GO\_0008236; F:serine-type peptidase activity; IEA.

DR GO; GO\_0005198; F:structural molecule activity; IEA.

DR GO; GO\_0016740; F:transferase activity; IEA.

DR GO; GO\_0005508; P:proteolysis and peptidolysis; IEA.

DR GO; GO\_0006350; P:transcription; IEA.

DR GO; GO\_0019079; P:viral genome replication; IEA.

DR GO; GO\_0019387; P:viral transformation; IEA.

DR InterPro; IPR009003; Cys Ser\_trypsin.

DR InterPro; IPR01410; DEAD.

DR InterPro; IPR002522; HCV\_capsid.

DR InterPro; IPR00521; HCV\_core.

DR InterPro; IPR00219; HCV\_env.

DR InterPro; IPR00231; HCV\_NS1.

DR InterPro; IPR00218; HCV\_NS2.

DR InterPro; IPR00745; HCV\_NS4a.

DR InterPro; IPR00490; HCV\_NS4b.

DR InterPro; IPR00268; HCV\_Nsa.

DR InterPro; IPR002166; HCV\_RRap.

DR InterPro; IPR001650; Helicase\_C.

DR InterPro; IPR004109; Peptidase\_C29.

DR InterPro; IPR007095; RNA\_pol\_DS\_ps.

DR InterPro; IPR007094; RNA\_poi\_F5V\_r.

DR Pfam; PF01543; HCV\_core; 1.

DR Pfam; PR01539; HCV\_env; 1.

DR Pfam; PR01560; HCV\_NS1; 1.

DR Pfam; PR01538; HCV\_NS2; 1.

DR Pfam; PR02907; HCV\_NS3; 1.

DR Pfam; PR01006; HCV\_NS4a; 1.

DR Pfam; PR0101; HCV\_NS4b; 1.

DR Pfam; PR01506; HCV\_NS5a; 1.

DR Pfam; PR00271; helicase\_C; 1.

DR Pfam; PR00998; viral RdRp; 1.

DR Pfam; PR18662; HCV\_NS1; 1.

SMART; SM00487; DDXDC; 1.

Coat protein; Glycoprotein; Nonstructural protein; Polypeptide; RNA-directed RNA polymerase; Transferase; Transmembrane.

KW Sequence 3010 AA; 326880 MW; EEDB40E6A050E CRC64;

**RESULT 9**

ID Q9J3H3 PRELIMINARY; PRT; 3010 AA.

AC Q9J3H3;

DT 01-OCT-2000 (TREMBrel. 15, Created)

DT 01-OCT-2000 (TREMBrel. 15, Last sequence update)

DT 01-OCT-2003 (TREMBrel. 25, Last annotation update)

DE Genome polyprotein.

OS Hepatitis C virus.

VIRUSES; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepatitis C virus.

OC HCV.

OX NCBI\_TAXID:1103;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=MD;9;

RT "Characteristics of hepatitis C viral genome associated with disease progression;"

Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.

- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MENA (BY SIMILARITY).

CC EMBL; AF207760; AAF65950.1; -.

DR Pfam; PR01542; HCV\_core; 1.

DR Pfam; PR01539; HCV\_env; 1.

DR Pfam; PR01560; HCV\_NS1; 1.

DR Pfam; PR01538; HCV\_NS2; 1.

DR Pfam; PR02907; HCV\_NS3; 1.

DR Pfam; PR01006; HCV\_NS4a; 1.

DR Pfam; PR0101; HCV\_NS4b; 1.

DR Pfam; PR01506; HCV\_NS5a; 1.

DR Pfam; PR00271; helicase\_C; 1.

DR Pfam; PR00998; viral RdRp; 1.

DR Pfam; PR18662; HCV\_NS1; 1.

SMART; SM00487; DDXDC; 1.

Coat protein; Glycoprotein; Nonstructural protein; Polypeptide; RNA-directed RNA polymerase; Transferase; Transmembrane.

KW Sequence 3010 AA; 326880 MW; EEDB40E6A050E CRC64;

Query Match 95.9%; Score 1005; DB 12; Length 3010;

Matches 358; Conservativeness 94.5%; Pred. No. 1-2e-15; Mismatches 9; Indels 0; Gaps 0;

Query 2 ULLTSPYYKULLARIWWQYLITRVEAHLLQWIPPLNRRGRDAILITCAHPELIFD 61

Db 828 LLLSPYYKULLARIWWQYLITRVEAHLLQWIPPLNRRGRDAILITCAHPELIFD 887

Query 62 ITKLIAILGPLMLQAGITRVPFVRAGQILRKVLAQMLVKAAGHIVQMAFVKLAALTGT 947

Db 886 ITKLIAILGPLMLQAGITRVPFVRAGQILRKVLAQMLVKAAGHIVQMAFVKLAALTGT 947

Query 12 YVYDHUTPQDWAAGLRILAVAEPVPTSDMEYKTTGADTAACGDIISGLPVSARRG 181

Db 948 YVYDHUTPQDWAAGLRILAVAEPVPTSDMEYKTTGADTAACGDIISGLPVSARRG 1007

Query 182 REILIGPANDFEGQWRLAPITYSQRGLGCITSLTGDKNQFGEVQVSTAQ 241

Db 1008 KEILIGPANSLEGQWRLAPITYSQRGLGCITSLTGDKNQFGEVQVSTAQ 1067

Query 242 SFLATCNGYCWTVFHGAGSKTLAGPKGPIOTQMYTNVDLGQAPQARSMPCTCGS 301

Db 1068 SFLATCNGYCWTVFHGAGSKTLAGPKGPIOTQMYTNVDLGQAPQARSMPCTCGS 1127

Query 302 SDLYLVTRHADVIPRRRGDSRSGLSPPVSYLKGSQSGPLCPSGHVGIFRAAVTR 361

Db 1128 SDLYLVTRHADVIPRRRGDSRSGLSPPVSYLKGSQSGPLCPSGHVGIFRAAVTR 1187

Query 362 GVAKAVDFPVESMETMR 1206

Db 1188 GVAKAVDFPVESMETMR 1206



DR InterPro; IPR004109; Peptidase\_C29.  
 DR InterPro; IPR007095; RNA pol DS PS.  
 DR Pfam; PF01543; HCV capsid; 1.  
 DR Pfam; PF01542; HCV core; 1.  
 DR Pfam; PF01539; HCV env; 1.  
 DR Pfam; PF01560; HCV\_NS1; 1.  
 DR Pfam; PF01538; HCV\_NS2; 1.  
 DR Pfam; PF01537; HCV\_NS3; 1.  
 DR Pfam; PF01006; HCV\_NS4; 1.  
 DR Pfam; PF01001; HCV\_NS4b; 1.  
 DR Pfam; PF01506; HCV\_NS5a; 1.  
 DR Pfam; PF01271; helicase\_C; 1.  
 DR Pfam; PF00598; viral\_RdRp; 1.  
 DR ProDom; PDB8602; HCV\_NS1; 1.  
 DR SMART; SM00487; DEXD\_C; 1.  
 DR PROSITE; PS00190; CTOCHROME\_C; 1.  
 DR Coat protein; Envelope protein; Glycoprotein; Nonstructural protein; Transferase; Transmembrane.  
 KW Polypeptide; RNA-directed RNA polymerase; RNA; Sequence  
 SQ SEQUENCE 3010 AA; 326909 MW; 5505C62EBODA0519 CRC64;

Query Match Best Local Similarity 95.9%; Score 1905; DB 12; Length 3010;  
 Matches 359; Conservatism 94.5%; Pred. No. 1; 2-e-151; Mismatches 11; Indels 0; Gaps 0;  
 Qy 1 ALITLSSYYKVLARLILWQIOLYTRTEAHQIOWQWIPPLNVRGGDRDILILTCVAVHPELIF 60  
 Db 827 ALITLSPYYKVFLARLILWQIOPITRAFAYLQWVPPLNVRGGDRDILILTCVAVHPELIF 886  
 Qy 61 DITKLIAIFGMLQLOGITKUPYFRAGGLIRACMLVKRAGGHVQMAFMKLAUTG 120  
 Db 887 DITKLIAIFGMLQLOGITKUPYFRAGGLIRACMLVKRAGGHVQMAFMKLAUTG 946

Qy 121 TVVYDHILPLOWAHAGRDIAVAVENWVFSMNEVKILTWDGDTAAQGDIISGLPSARR 180  
 Db 947 TVVYDHILPLOWAHAGRDIAVAVENWVFSMNEVKILTWDGDTAAQGDIISGLPSARR 1005

Qy 181 GRBLLGPADNFEGQGNLLATITAYSOOTRNLIGCTTSTLGDRDKNQVEGVQVSTAT 240  
 Db 1067 QSPFLATCINGVWTVFHGASKTLAGPKGPITQMYINVDQDQVWGPGRSMPTCTG 1126

Db 1007 GRBLLGPADNFEGQGNLLATITAYSOOTRNLIGCTTSTLGDRDKNQVEGVQVSTAT 1066

Qy 241 QSPFLATCINGVWTVFHGASKTLAGPKGPITQMYINVDQDQVWGPGRSMPTCTG 300

Db 301 SDLYJYJTRHAYVIPVRRGGDRGSLSLSPRPSYLKSSGCPLLCPSGHANGIFAAVT 360  
 Qy 1127 SDPLYJYTRHAYVIPVRRGGDRGSLSLSPRPSYLKSSGCPLLCPSGHANGIFAAVT 1186

Qy 361 RSYKAKVADIFPESMETMR 380  
 Db 1187 RSYKAKVADIFPESMETMR 1206

RESULT 11  
 Q81755 PRELIMINARY; PRT; 1186 AA.  
 ID Q81755  
 AC Q81755  
 DT 01-NOV-1996 (TREMBrel\_01, Created)  
 DT 01-NOV-1996 (TREMBrel\_01, Last sequence update)  
 DT 01-OCT-2003 (TREMBrel\_25, Last annotation update)  
 DE Polyprotein (Fragment).  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Herpesvirus.  
 OX NCBI\_TAXID=1103;  
 RN [1]  
 RP SSQUENCE FROM N.A.  
 MEDLINE=9108550; PubMed=2175903;  
 RA Kato N., Hijikata M., Ootsuyama Y., Nakayama M., Ohkoshi S.,  
 RA Sugimura T., Shimotohno K.; "Molecular cloning of the human hepatitis C virus genome from Japanese patients with non-A, non-B hepatitis.";

RL PROC. Natl. Acad. Sci. U.S.A. 87:9524-9528(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=9204440; PubMed=1658196;  
 RA Okamoto H., Okada S., Sugiyama Y., Kurai K., Iizuka H., Machida A.,  
 RA Miyakawa Y., Mayumi M.,  
 RT "Nucleotide sequences of the genomic RNA of hepatitis C virus isolated from a human carrier: comparison with reported isolates for conserved and divergent regions.";  
 RT J. Gen. Virol. 72:2697-2704(1991).  
 RL RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=9114698; PubMed=1847440;  
 RA Takemoto A., Mori C., Manabe S., Murakami S., Fujita J., Onishi E.,  
 RA Andon T., Yoshida I., Okanya H.; "The structure and organization of the Hepatitis C virus genome isolated from human carriers.";  
 RT J. Virol. 65:1105-1113(1991).  
 RP [4]  
 SEQUENCE FROM N.A.  
 RX MEDLINE=91172826; PubMed=1848704;  
 RA Choo Q.-L., Richman K., Han J.H., Berger K., Lee C., Dong C.,  
 RA Gallegos M., Coit D., Medina-Selby A., Barr P.J., Weiner A.,  
 RA Bradley D.W., Kuo G., Houghton M.,  
 RT "Genetic Organization and diversity of the hepatitis C virus.";  
 Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).  
 RL RN [5]  
 SEQUENCE FROM N.A.  
 RX MEDLINE=91172826; PubMed=1848704;  
 RA Chen P., Li M., Tai K., Liu P., Lin C., Chen D.;  
 RA Fukuda S., Tsuda F., Mishiro S.; "The Taiwanese hepatitis C virus genome: Sequence determination and mapping the 5' termini of viral genomic and antigenomic RNA.";  
 RT Virology 188:102-113(1992).  
 RL RN [6]  
 SEQUENCE FROM N.A.  
 RX MEDLINE=92230506; PubMed=1314459;  
 RA Okamoto H., Kurai K., Okada S., Yamamoto K., Iizuka H., Tanaka T.,  
 RA Fukuda S., Tsuda F., Mishiro S.; "The full-length sequence of a hepatitis C virus genome having poor homology to reported isolates: Comparative study of four distinct genotypes.";  
 RT Virology 188:331-341(1992).  
 RL RN [7]  
 SEQUENCE FROM N.A.  
 RX MEDLINE=93323008; PubMed=8392606;  
 RA Hijikata M., Mizushima H., Akagi T., Mori S., Kakiuchi N., Kato N.,  
 RA Tanaka T., Kimura K., Shimotohno K.; "Two distinct proteolytic activities required for the processing of a putative nonstructural precursor protein of hepatitis C virus.";  
 RT J. Virol. 67:4665-4675(1993).  
 RL RN [8]  
 SEQUENCE FROM N.A.  
 RA Hijikata M.; "Submitted (DEC-1993) to the EMBL/GenBank/DDBJ databases."  
 RL EMBL; D11397; BRA0975.1; -.  
 DR PIR; A61196; A61195.  
 DR PIR; PS0329; PS0328.  
 DR PDB; 1DXP; 2B-MAR-02.  
 DR GO; GO:0005524; F-ATP binding; IEA.  
 DR GO; GO:0008025; F-ATP dependent helicase activity; IEA.  
 DR GO; GO:0016787; F:Hydrolease activity; IEA.  
 DR GO; GO:0003676; F:nucleic acid binding; IEA.  
 DR GO; GO:0008235; F:serine-type peptidase activity; IEA.  
 DR GO; GO:0006508; F:proteolysis and peptidolysis; IEA.  
 DR GO; GO:0019087; F:viral transformation; IEA.  
 DR InterPro; IPR009003; Cys\_Ser\_trypsin.  
 DR InterPro; IPR01410; DEAD.  
 DR InterPro; IPR002518; HCV\_NS2.  
 DR InterPro; IPR000745; HCV\_NS4.  
 DR InterPro; IPR001490; HCV\_NS4b.  
 DR InterPro; IPR004109; Helicase\_C.  
 DR Pfam; PF01538; HCV\_NS2; 1.





"Characteristics of hepatitis C viral genome associated with disease progression.", Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.

!- SUBMIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA. (BY SIMILARITY).

EMBL; AR207754; AAH65944.1; - .

FIR; A61196; A61196.

DR; PIR; PQ0246; PQ0246.

DR; PIR; PS0329; PS0329.

HSSP; P2663; IJXP2.

DR; GO; GO\_001602; C:integral to membrane; IEA.

DR; GO; GO\_0019028; C:viral capsid; IEA.

DR; GO; GO\_001931; C:viral envelope; IEA.

DR; GO; GO\_0005524; F:ATP binding; IEA.

DR; GO; GO\_0008026; F:ATP dependent helicase activity; IEA.

DR; GO; GO\_0005489; F:electron transporter activity; IEA.

DR; GO; GO\_003723; F:RNA binding; IEA.

DR; GO; GO\_000368; F:RNA-directed RNA polymerase activity; IEA.

DR; GO; GO\_0008236; F:serine-type peptidase activity; IEA.

DR; GO; GO\_0005398; F:structural molecule activity; IEA.

DR; GO; GO\_0016740; F:transferase activity; IEA.

DR; GO; GO\_0006118; P:electron transport; IEA.

DR; GO; GO\_0006508; P:proteolysis and peptidolysis; IEA.

DR; GO; GO\_0006350; P:transcription; IEA.

DR; GO; GO\_0019079; P:viral genome replication; IEA.

DR; InterPro; IPR009003; Cys\_Ser\_Trypsin.

InterPro; IPR00345; Cytc\_heme\_BS.

InterPro; IPR001410; DEAD.

InterPro; IPR0522; HCV\_capsid.

InterPro; IPR002521; HCV core.

InterPro; IPR00519; HCV\_env.

InterPro; IPR0253; HCV\_NS1.

InterPro; IPR002518; HCV\_NS2.

InterPro; IPR00745; HCV\_NS4.

InterPro; IPR01490; HCV\_NS4b.

InterPro; IPR002868; HCV\_N5a.

InterPro; IPR00166; HCV\_RdRp.

InterPro; IPR01650; Helicase\_C.

InterPro; IPR0410; Peptidase\_C29.

InterPro; IPR007095; RNA\_pol\_BS\_Ps.

InterPro; IPR007094; RNA\_Pol\_PsVir.

Pfam; PF01543; HCV\_capsid; 1.

Pfam; PF01542; HCV\_core; 1.

Pfam; PF00539; HCV\_env; 1.

Pfam; PF01560; HCV\_NS1; 1.

Pfam; PF00538; HCV\_NS2; 1.

Pfam; PF00907; HCV\_NS3; 1.

Pfam; PF01006; HCV\_NS4; 1.

Pfam; PF0001; HCV\_NS4b; 1.

Pfam; PF01506; HCV\_N5a; 1.

Pfam; PF00999; Viral\_RdRp; 1.

ProDom; PF18602; HCV\_NS1; 1.

SMART; SM00487; DEXD\_C; 1.

DR; PROSITE; PS00190; CYTOCHROME\_C; 1.

KW; Coat\_protein; Envelope\_protein; Glycoprotein; Nonstructural\_protein; Polypeptide; RNA-directed RNA\_polymerase; Transferase; Transmembrane.

SEQUENCE 3010 AA; 32984 MW; AF12CC00E0B08B078 CRC64;

Query Match 95.7%; Score 1902; DB 12; Length 3010;

Best Local Similarity 93.7%; Pire No. 2.2e-15;

Matches 356; Conservative 11; Mismatches 13; Indels 0; Gaps 0;

Qy 1 ALITSPYKVYLARLILWQQLYRVEHLQWNPPLNVRGDRDAILLTCAVHPFLF 60

Db 827 ALITSPYKVYLARLILWQQLYRVEHLQWNPPLNVRGDRDAILLTCAVHPFLF 886

61 DITKULLAIFLGPLMLQAGITKIVYFVRAGLIRACMVLRAAGHIVOMAFNKLALTG 120

Qy 887 DITKULLAIFLGPLMLQAGITKIVYFVRAGLIRACMVLRAAGHIVOMAFNKLALTG 946

RESULT 15

Q99AU2 ID Q99AU2 PRELIMINARY PRT; -310 AA.

AC Q99AU2; DT 01-JUN-2001 (TREMBrel. 17, Created)

DT 01-JUN-2001 (TREMBrel. 17, Last sequence update)

DT 01-OCT-2003 (TREMBrel. 25, Last annotation update)

DE Genome polyprotein.

OG Hepatitis C virus type 1b.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.

OX NCBI\_TaxID=31647;

[1]

RN RPP

RC STRAIN=chimera of HCV-BK;

RA Thomson M., Nascimbene M., Gonzales S., Murthy K., Rehermann B., Liang J.,

RT "Analyses of viral sequences and virus-specific immune responses during serial passage of an infectious hepatitis C virus serotype 1b clone in chimpanzees";

RT Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.

CC -!- SUBMIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA. (BY SIMILARITY).

EMBL; AR33324; AAK0509.1; - .

DR; PIR; A61196; A61196.

DR; PIR; PQ0246; PQ0246.

DR; PIR; PQ0804; PQ0804.

DR; PIR; PS0329; PS0329.

HSSP; P2663; IN33.

DR; GO; GO\_001602; C:integral to membrane; IEA.

DR; GO; GO\_0019028; C:viral capsid; IEA.

DR; GO; GO\_001931; C:viral envelope; IEA.

DR; GO; GO\_0005524; F:ATP binding; IEA.

DR; GO; GO\_000026; F:ATP dependent helicase activity; IEA.

DR; GO; GO\_0005489; F:electron transporter activity; IEA.

DR; GO; GO\_0003350; P:transcription; IEA.

DR; GO; GO\_0019079; P:viral genome replication; IEA.

DR; GO; GO\_0003968; F:RNA-directed RNA polymerase activity; IEA.

DR; GO; GO\_0008236; F:serine-type peptidase activity; IEA.

DR; GO; GO\_000519; F:structural molecule activity; IEA.

DR; GO; GO\_0016740; F:transferase activity; IEA.

DR; GO; GO\_0006118; P:electron transport; IEA.

DR; GO; GO\_0006508; P:proteolysis and peptidolysis; IEA.

DR; GO; GO\_0003350; P:transcription; IEA.

DR; GO; GO\_0019079; P:viral genome replication; IEA.

DR; GO; GO\_0019087; P:viral transformation; IEA.

DR; InterPro; IPR009003; Cys\_Ser\_Trypsin.

InterPro; IPR00345; Cytc\_heme\_BS.

InterPro; IPR001410; DEAD.

DR; InterPro; IPR002522; HCV\_capsid.

DR InterPro; IPR002521; HCV\_core.  
 DR InterPro; IPR002513; HCV\_env.  
 DR InterPro; IPR002531; HCV\_NS1.  
 DR InterPro; IPR002518; HCV\_NS2.  
 DR InterPro; IPR000743; HCV\_NS4.  
 DR InterPro; IPR001490; HCV\_NS4D.  
 DR InterPro; IPR002868; HCV\_NS5.  
 DR InterPro; IPR002166; HCV\_RdR.  
 DR InterPro; IPR004109; Peptidase\_C29.  
 DR InterPro; IPR007095; RNA\_Pol\_Ds\_ps.  
 DR InterPro; IPR007094; RNA\_Pol\_PsVir.  
 DR Pfam; PF01543; HCV\_gapsid; 1.  
 DR Pfam; PF01542; HCV\_core; 1.  
 DR Pfam; PF01539; HCV\_env; 1.  
 DR Pfam; PF01560; HCV\_NS1; 1.  
 DR Pfam; PF01538; HCV\_NS2; 1.  
 DR Pfam; PF02907; HCV\_NS3; 1.  
 DR Pfam; PF01005; HCV\_NS4a; 1.  
 DR Pfam; PF01001; HCV\_NS4b; 1.  
 DR Pfam; PF01506; HCV\_NS5a; 1.  
 DR Pfam; PF00998; Viral\_RNP; 1.  
 DR Prodrom; PDI18662; HCV\_NS1; 1.  
 DR SMART; SM00487; DEXIC; 1.  
 DR PROSITE; PS00190; CYTOCHROME\_C; 1.  
 KW Coat\_Protein; Envelope\_protein; Glycoprotein; Nonstructural\_protein;  
 KW Polypeptide; RNA-directed\_RNA\_polymerase; Transferase; Transmembrane.  
 SQ Sequence 3010 AA; 32707 MW; 053BA6530AB35 CPC64;

Query Match 95.7%; Score 1902; DB 12; Length 3010;  
 Best Local Similarity 94.2%; Freq. No. 2.2e-151; Mismatches 10; Indels 0; Gaps 0;  
 Matches 357; Conservative 12; Mismatches 10; Indels 0; Gaps 0;

```

QY 2 LITLSPYKYLARLIWQQLIRVEAHQWWPLNVRGGRDAILITCAVHPFLFD 61
Db 828 LITLSSYYKYLARLIWQQLIRVEAHQWWPLNVRGGRDAILITCAVHPFLFD 887
QY 62 ITKLILIAIFGELMVHQAGITKVPPVRAQGLIRACMLVKAGGHYVOMAFMKLAITGT 121
Db 888 ITKLILIAIFGELMVHQAGITKVPPVRAQGLIRACMLVKAGGHYVOMAFMKLAITGT 947
QY 122 YVYDHHTPLQWAHAGIRDIAVAVPVPITSDMEYKTTWGAATACGDIISGILPVSARG 181
Db 948 YVNHHTPLRWAHAGIRDIAVAVPVPVSDFMETKLTWGAATACGDIILPVSARG 1007
QY 182 REBLIGGPADNEQGQMRLLAPITASQQTGGCITTSITGDKDKNQEGEYQVSTAQ 241
Db 1008 KEILLGPADSELEGRGRWLLAFTIASQQTGGCITTSITGDKDKNQEGEYQVSTAQ 1067
QY 242 SFLATCNGVWTWVHGAGSKTLAGPKGITQMTNTVVDIYVWQAPPGARNTPCTGS 301
Db 1068 SFLATCNGVWTWVHGAGSKTLAGPKGITQMTNTVVDIYVWQAPPGARNTPCTGS 1127
QY 302 SDLYVWTRHADWVTPVRRGSRGSLSLSPPVSYKGSSCGPILCPSGHAWGIFRAAVCTR 361
Db 1128 SDLYVWTRHADWVTPVRRGSRGSLSLSPPVSYLKGSCLCPSGHAWGIFRAAVCTR 1187
QY 362 GYAKAIDFIPESMETM 380
Db 1188 GVAKAVDFPVPEM 1206

```

Search completed: May 6, 2004, 09:35:47  
 Job time : 36.55407 secs



Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:25:16 ; Search time 14.914 Seconds  
 (Without alignments)  
 1315.364 Million cell updates/sec

Title: US-10-650-585-12  
 Perfect score: 1987  
 Sequence: 1 ALITLSPYKVLLRLIWL.....RGWAKAVDFIPVESMETMR 380

Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
 Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries

Database : Issued Patents AA:  
 1: /cgn2\_6/ptodata/2/iaa/5A\_COMB\_pep: \*  
 2: /cgn2\_6/ptodata/2/iaa/5B\_COMB\_pep: \*  
 3: /cgn2\_6/ptodata/2/iaa/6A\_COMB\_pep: \*  
 4: /cgn2\_6/ptodata/2/iaa/6B\_COMB\_pep: \*  
 5: /cgn2\_6/ptodata/2/iaa/PCUS\_COMB\_pep: \*  
 6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match Length | DB ID | Description         |
|------------|-------|--------------------|-------|---------------------|
| 1          | 1897  | 95.0               | 2201  | 4 US-09-539-601-6   |
| 2          | 1887  | 95.0               | 2201  | 4 US-09-539-601-15  |
| 3          | 1887  | 95.0               | 3010  | 4 US-09-539-601-3   |
| 4          | 1887  | 95.0               | 3010  | 4 US-09-539-601-21  |
| 5          | 1887  | 94.7               | 3010  | 4 US-09-539-601-27  |
| 6          | 1881  | 94.7               | 1692  | 3 US-09-243-934-4   |
| 7          | 1881  | 94.7               | 1692  | 4 US-09-919-901-4   |
| 8          | 1881  | 94.7               | 2307  | 3 US-09-263-935-2   |
| 9          | 1881  | 94.7               | 2307  | 4 US-09-919-901-2   |
| 10         | 1880  | 94.6               | 3010  | 4 US-09-539-601-33  |
| 11         | 1879  | 94.5               | 1692  | 3 US-09-263-931-11  |
| 12         | 1878  | 94.5               | 1692  | 4 US-09-919-901-11  |
| 13         | 1878  | 94.5               | 2307  | 3 US-09-263-931-9   |
| 14         | 1869  | 94.1               | 1692  | 3 US-09-263-931-18  |
| 15         | 1869  | 94.1               | 1692  | 4 US-09-919-901-18  |
| 16         | 1869  | 94.1               | 1692  | 4 US-09-919-901-16  |
| 17         | 1869  | 94.1               | 2307  | 3 US-09-263-931-16  |
| 18         | 1869  | 94.1               | 2307  | 4 US-09-919-901-16  |
| 19         | 1869  | 94.1               | 3010  | 3 US-09-914-416-3   |
| 20         | 1823  | 91.7               | 2013  | 2 US-08-324-977-12  |
| 21         | 1823  | 91.7               | 2013  | 2 US-08-384-616-12  |
| 22         | 1823  | 91.7               | 2013  | 2 US-08-904-686A-12 |
| 23         | 1823  | 91.7               | 2013  | 3 US-09-318-855-12  |
| 24         | 1823  | 91.7               | 2201  | 4 US-08-952-98A-2   |
| 25         | 1823  | 91.7               | 2620  | 1 US-08-324-977-32  |
| 26         | 1823  | 91.7               | 2620  | 2 US-08-384-616-32  |
| 27         | 1823  | 91.7               | 2620  | 2 US-08-904-686A-32 |

ALIGNMENTS

```

RESULT 1
US-09-539-601-6
; Sequence 6, Application US/09539601C
; Patent No. 6630343
; GENERAL INFORMATION:
; APPLICANT: Bartenschlager, Ralf FW
; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System
; FILE REFERENCE: all sequences
; CURRENT APPLICATION NUMBER: US/09/539,601C
; CURRENT FILING DATE: 2001-08-30
; EARLIER FILING DATE: 1993-04-03
; BARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO: 6
; LENGTH: 2201
; TYPE: PRT
; ORGANISM: Hepatitis C virus
; US-09-539-601-6

Query Match Similarity 95.0%; Score 1897; DB 4; Length 2201;
Best Local Similarity 93.4%; Pred. No. 3.8e-181; Mismatches 12; Indels 0; Gaps 0;
Matches 354; Conservancy 13; N mismatches 12; Sequence 6, Appli
Qy 2 LITLSPYKVLLRLIWLWYLQLRIVAEQLQWVPPLNRRGRDAILITCAVHPFLFD 61
Db 19 LITLSPHYKFLARLILWWLYQFTRAAHLQWVPPLNRRGRDAILITCAVHPFLFD 78
Qy 62 ITKLIAIFGLPMVQAGITKVPFVRAGLIRACMVLKAAGCHYVQAFMKLAIGT 121
Db 79 ITKLIAIFGLPMVQAGITKVPFVRAGLIRACMVLKAAGCHYVQAFMKLAIGT 138
Qy 122 YVDHHTPLQDWAHGLRDLAVAVPEVPISDMEVKILWQGDTACGDISGIPVSARRG 181
Db 139 YVDHHTPLQDWAHGLRDLAVAVPEVPISDMEVKILWQGDTACGDISGIPVSARRG 198
Qy 182 REILGPADNPEGOQWRLAPITASQQRGLGCITSLGRDKNQVQGBVQNSTAQ 241
Db 199 REILGPADNPEGOQWRLAPITASQQRGLGCITSLGRDKNQVQGBVQNSTAQ 258
Qy 242 SFLATCNGVCTWTFHGAGSKTLAGPKPQTOMYNTNDLGVQAPPARSMTPTCS 301
Db 259 SFLATCNGVCTWTFHGAGSKTLAGPKPQTOMYNTNDLGVQAPPARSMTPTCS 318
Qy 302 SDLYVUTRHDVYPRRGSGRSLSPRVSYLGSSGCPCLPSGHAVGIFRAVCTR 361
Db 319 SDLYVUTRHDVYPRRGSGRSLSPRVSYLGSSGCPCLPSGHAVGIFRAVCTR 378
Qy 362 GVAKAVDFIPVESMETMR 380

```

Db 379 GVAKAVDFVVPVESMETTMR 397

; ORGANISM: Hepatitis C virus  
US-09-539-601-3

US-09-539-601-15  
; Sequence 15, Application US/09539601C  
; Patent No. 663043  
; GENERAL INFORMATION:  
; APPLICANT: Bartenschlager, Ralf FW  
; FILE REFERENCE: all sequences  
; CURRENT APPLICATION NUMBER: US/09/539,601C  
; CURRENT FILING DATE: 2001-08-30  
; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY  
; EARLIER FILING DATE: 1999-04-03  
; NUMBER OF SEQ ID NOS: 51  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 15  
; LENGTH: 2201  
; TYPE: PRT  
; ORGANISM: Hepatitis C virus  
; US-09-539-601-15

Query Match 95.0%; Score 1887; DB 4; Length 2201;  
Best Local Similarity 93.4%; Pred. No. 3.8e-181; Mismatches 13; Matches 354; Conservative 13; Indels 0; Gaps 0;  
Matches 354; Conservative 13; Mismatches 12; Indels 0; Gaps 0;

Qy 2 LITLSPYKVLARLTIWMLQQLITRVEAHQWVPPNLRGGDAILITCAVHPELIED 61  
Db 828 LITLSPYKVLARLTIWMLQQLITRVEAHQWVPPNLRGGDAILITCAVHPELIFT 887

Qy 62 ITKLLAIFGPMVLOGLGTTKVPFVRAQGLIRACMLVRKAAGGHYVOMAFMKAALGT 121  
Db 888 ITKLLAIFGPMVLOGLGTTKVPFVRAQGLIRACMLVRKAAGGHYVOMAFMKAALGT 947

Qy 122 YVYDHLTPLQDWAHAGRLDAVAVEPVVFSDMETKIVTGADTAACGDIISGIPVSARRG 181  
Db 1008 YVYDHLTPLQDWAHAGRLDAVAVEPVVFSDMETKIVTGADTAACGDIISGIPVSARRG 1007

Qy 948 YVYDHLTPLQDWAHAGRLDAVAVEPVVFSDMETKIVTGADTAACGDIISGIPVSARRG 1007

Qy 182 RBLIGPADNTRGQMRLLAPITAVSQTRGGLGCTTSITGRDNQKRNQEVGEOVUSTAQ 241  
Db 1068 RBLIGPADNTRGQMRLLAPITAVSQTRGGLGCTTSITGRDNQKRNQEVGEOVUSTAQ 1067

Qy 302 SDLYLVTRHADVPIVRRGDSRSGLSISPRPVSYLKGSQSGGPCLCPGHAVGIFRAAVCTR 361  
Db 1128 SDLYLVTRHADVPIVRRGDSRSGLSISPRPVSYLKGSQSGGPCLCPGHAVGIFRAAVCTR 1187

Qy 362 GVAKADFIPVESMETMR 380  
Db 1188 GVAKADFIPVESMETMR 1206

Qy 139 YVYDHLTPLQDWAHAGRLDAVAVEPVVFSDMETKIVTGADTAACGDIISGIPVSARRG 198

Qy 182 REILIGPADNTRGQMRLLAPITAVSQTRGGLGCTTSITGRDNQKRNQEVGEOVUSTAQ 241

Qy 199 REIHGPALSLEGQWRLLAPITAVSQTRGGLGCTTSITGRDNQKRNQEVGEOVUSTAQ 258

Qy 242 SFLATCNGVWTVFHGAGSKTLAGPKGKPTQMTWVQDODLGQAPPARSILPCTCGS 301

Qy 259 SFLATCNGVWTVFHGAGSKTLAGPKGKPTQMTWVQDODLGQAPPARSILPCTCGS 318

Qy 302 SDLYLVTRHADVPIVRRGDSRSGLSISPRPVSYLKGSQSGGPCLCPGHAVGIFRAAVCTR 361

Qy 319 SDLYLVTRHADVPIVRRGDSRSGLSISPRPVSYLKGSQSGGPCLCPGHAVGIFRAAVCTR 378

Qy 362 GVAKADFIPVESMETMR 380

Qy 379 GVAKADFIPVESMETMR 397

RESULT 3  
US-09-539-601-3  
; Sequence 3, Application US/09539601C  
; Patent No. 663043  
; GENERAL INFORMATION:  
; APPLICANT: Bartenschlager, Ralf FW  
; FILE REFERENCE: all sequences  
; CURRENT APPLICATION NUMBER: US/09/539,601C  
; CURRENT FILING DATE: 2001-08-30  
; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY  
; EARLIER FILING DATE: 1999-04-03  
; NUMBER OF SEQ ID NOS: 51  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 21  
; LENGTH: 3010  
; TYPE: PRT  
; ORGANISM: Hepatitis C virus  
; US-09-539-601-3

Query Match 95.0%; Score 1887; DB 4; Length 3010;  
Best Local Similarity 93.4%; Pred. No. 6.1e-181; Mismatches 13; Matches 354; Conservative 13; Indels 0; Gaps 0;

Qy 2 LITLSPYKVLARLTIWMLQQLITRVEAHQWVPPNLRGGDAILITCAVHPELIED 61  
Db 828 LITLSPYKVLARLTIWMLQQLITRVEAHQWVPPNLRGGDAILITCAVHPELIFT 887

Qy 62 ITKLLAIFGPMVLOGLGTTKVPFVRAQGLIRACMLVRKAAGGHYVOMAFMKAALGT 121  
Db 888 ITKLLAIFGPMVLOGLGTTKVPFVRAQGLIRACMLVRKAAGGHYVOMAFMKAALGT 947

Qy 122 YVYDHLTPLQDWAHAGRLDAVAVEPVVFSDMETKIVTGADTAACGDIISGIPVSARRG 181  
Db 948 YVYDHLTPLQDWAHAGRLDAVAVEPVVFSDMETKIVTGADTAACGDIISGIPVSARRG 1007

Qy 182 REILIGPADNTRGQMRLLAPITAVSQTRGGLGCTTSITGRDNQKRNQEVGEOVUSTAQ 241

RESULT 5  
 US-09-539-601-27  
 ; Sequence 27, Application US/09539601C  
 ; Patent No. 6330343  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Bartenschlager, Ralf FW  
 ; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System  
 ; FILE REFERENCE: 0125-0005A  
 ; CURRENT APPLICATION NUMBER: US/09/263, 933  
 ; CURRENT FILING DATE: 1999-03-08  
 ; EARLIER APPLICATION NUMBER: 09/129, 611  
 ; EARLIER FILING DATE: 1998-08-05  
 ; NUMBER OF SEQ ID NOS: 33  
 ; SOFTWARE: Patentin Ver. 2.0  
 ; SEQ ID NO: 4  
 ; LENGTH: 1692  
 ; TYPE: PRT  
 ; ORGANISM: Hepatitis C virus  
 ; US-09-539-601-27  
 Query Match 95.0%; Score 1887; DB 4; Length 3010;  
 Best Local Similarity 93.4%; Pred. No. 6.1e-81; Mismatches 353; Conservatve 12; Indels 0; Gaps 0;  
 Matches 354; Conservative 13; Mismatches 12; Indels 0; Gaps 0;  
 Matches 353; Conservatve 12; Mismatches 14; Indels 0; Gaps 0;  
 Query 2 LILTSPYVYKVLARLIWQIYITRVAHLQWIPPLNVRGGDAITILTCAVHPELIFD 61  
 Db 828 LILTSPYVYKVLARLIWQIYITRVAHLQWIPPLNVRGGDAITILTCAVHPELIFT 887  
 Query 62 ITKLLIAITGPLMYLQAGITKRVYFVRAGLITRACMLVRAKGHHYQMAFMKLAACTGT 121  
 Db 167 ITKLLIAITGPLMYLQAGITKRVYFVRAGLITRACMLVRAKGHHYQMAFMKLGAITGT 226  
 Query 122 YVIDLTPPQDWHAHGDLAVAVEPVIFSDMVKITIWGADTAACGDIISGLPVSRG 181  
 Db 227 YIVNHLTPRLWDWAHAGLDRDLAGLAVAPEPVIFSDMVKITIWGADTAACGDIILGLPVSRG 286  
 Query 182 REILGPADNFEGQWRLLAPIAYSQOTRGIGCITSLTDKNQVEGEVQVSTATQ 241  
 Db 287 KEILLGPASLEGWRLLAPIAYSQOTRGIGCITSLTDKNQVEGEVQVSTATQ 346  
 Query 242 SFLATCVNGVCWTFHGAGSKTLAGPKGPIOTOMTNQDQJWQAPGASMTPTCGS 301  
 Db 1008 REILHPADSLAQGWRLAPIAYSQOTRGIGCITSLTDKNQVEGEVQVSTATQ 1067  
 Query 242 SFLATCVNGVCWTFHGAGSKTLAGPKGPIOTOMTNQDQJWQAPGASMTPTCGS 301  
 Db 1068 SFLATCVNGVCWTFHGAGSKTLAGPKGPIOTOMTNQDQJWQAPGASMTPTCGS 1127  
 Query 302 SDLYLVTRADVIFPVRREGDSRSLSLSPRVSYKGSGGPPLCPSCAVGIRAAVCTR 361  
 Db 1128 SDLYLVTRADVIFPVRREGDSRSLSLSPRVSYKGSGGPPLCPSCAVGIRAAVCTR 1187  
 Query 362 GVAKAVDFPVESMETMR 380  
 Db 1188 GVAKAVDFPVESMETMR 1206

RESULT 6  
 US-09-263-933-4  
 ; Sequence 4, Application US/09263933  
 ; Patent No. 6280340  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Potts, Karen E.  
 ; APPLICANT: Jackson, Roberta L.  
 ; APPLICANT: Patick, Amy K.  
 ; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
 ; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE  
 ; FILE REFERENCE: 0125-0005A  
 ; CURRENT APPLICATION NUMBER: US/09/919, 901  
 ; CURRENT FILING DATE: 2001-08-02  
 ; PRIOR APPLICATION NUMBER: 09/263, 933  
 ; PRIOR FILING DATE: 1999-02-08  
 ; PRIOR APPLICATION NUMBER: 09/129, 611  
 ; PRIOR FILING DATE: 1998-08-05  
 ; NUMBER OF SEQ ID NOS: 33  
 ; SOFTWARE: Patentin Ver. 2.0  
 ; SEQ ID NO: 4  
 ; LENGTH: 1692

RESULT 7  
 US-09-919-901-4  
 ; Sequence 4, Application US/09919901  
 ; Patent No. 6599738  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Potts, Karen E.  
 ; APPLICANT: Jackson, Roberta L.  
 ; APPLICANT: Patick, Amy K.  
 ; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
 ; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE  
 ; FILE REFERENCE: 0125-0005A  
 ; CURRENT APPLICATION NUMBER: US/09/919, 901  
 ; CURRENT FILING DATE: 2001-08-02  
 ; PRIOR APPLICATION NUMBER: 09/263, 933  
 ; PRIOR FILING DATE: 1999-02-08  
 ; PRIOR APPLICATION NUMBER: 09/129, 611  
 ; PRIOR FILING DATE: 1998-08-05  
 ; NUMBER OF SEQ ID NOS: 33  
 ; SOFTWARE: Patentin Ver. 2.0  
 ; SEQ ID NO: 4  
 ; LENGTH: 1692

; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE: OTHER INFORMATION: :  
; US-09-919-901-4

Query Match 94.7%; Score 1881; DB 4; Length 1692;  
Best Local Similarity 93.1%; Pred. No. 1..e-180; Mismatches 14; Indels 0; Gaps 0;  
Matches 353; Conservative 12; Mismatches 14; Indels 0; Gaps 0;

QY 2 LITLSPYYKVLLARLILWLOQITRVAHLQWIPPLNVRGRDAILITCAVHPELF 61  
Db 107 LITLSPYYKVFLARLILWLOQITRVAHLWIPPLNARGRDAILITCAVHPELF 166

QY 62 ITKLLAIFGPMVLTQAGITKPYFRAQGLIRACMVKAGGHVQMAFMKLAUTG 121  
Db 167 ITKLLAIFGPMVLTQAGITKPYFRAQGLIRACMVKAGGHVQMAFMKLAUTG 226

QY 122 YVDLHLPQLQWAHAGRLDVLAVAEPVTFSDMEVKITWGADTAACGDIISGLPVSARG 181  
Db 227 YVNHLTLPDQWAHAGRLDVLAVAEPVTFSDMEVKITWGADTAACGDIISGLPVSARG 286

QY 182 REILGPADNFEGQWRLLAPITASQOTRGIGCITSITGRDKNOVEGEVQVSTATQ 241  
Db 287 KEILGPADSLERGWRLLAPITASQOTRGIGCITSITGRDKNOVEGEVQVSTATQ 346

QY 242 SPLATCNGVCTVYHGASKTKLAGPKPITOMYTNDQDVGWQAPPGRASMTCTCGS 301  
Db 347 SPLATCNGVCTVYHGASKTKLAGPKPITOMYTNDQDVGWQAPPGRASMTCTCGS 406

QY 302 SDLYLVTRHADVTPVRGDSRSLSRSPVSYLKGSGGPLCPSGHAYGIFRAVCTR 361  
Db 407 SDLYLVTRHADVTPVRGDSRSLSRSPVSYLKGSGGPLCPSGHAYGIFRAVCTR 466

QY 362 GVAKAVDFPVESMETMR 380  
Db 467 GVAKAVDFPVESMETMR 485

RESULT 8  
US-09-263-933-2  
; Sequence 2, Application US/09263933  
; Patent No. 6280940  
; GENERAL INFORMATION:  
; APPLICANT: Potts, Karen E.  
; APPLICANT: Jackson, Roberta L.  
; APPLICANT: Potts, Karen E.  
; APPLICANT: Jackson, Roberta L.  
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
; CURRENT APPLICATION NUMBER: US/09/919, 901  
; CURRENT FILING DATE: 2001-08-02  
; PRIOR APPLICATION NUMBER: 09/263, 933  
; PRIOR FILING DATE: 1999-02-08  
; PRIOR APPLICATION NUMBER: 09/129, 611  
; PRIOR FILING DATE: 1998-08-05  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 2307  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE: OTHER INFORMATION: :  
; US-09-919-901-2

Query Match 94.7%; Score 1881; DB 4; Length 2307;  
Best Local Similarity 93.1%; Pred. No. 1..7e-180; Mismatches 14; Indels 0; Gaps 0;  
Matches 353; Conservative 12; Mismatches 14; Indels 0; Gaps 0;

QY 2 LITLSPYYKVLLARLILWLOQITRVAHLQWIPPLNVRGRDAILITCAVHPELF 61  
Db 139 LITLSPYYKVFLARLILWLOQITRVAHLWIPPLNARGRDAILITCAVHPELF 258

QY 62 ITKLLAIFGPMVLTQAGITKPYFRAQGLIRACMVKAGGHVQMAFMKLAUTG 121  
Db 259 ITKLLAIFGPMVLTQAGITKPYFRAQGLIRACMVKAGGHVQMAFMKLAUTG 318

Db 122 YVDLHLPQLQWAHAGRLDVLAVAEPVTFSDMEVKITWGADTAACGDIISGLPVSARG 181  
Db 319 YVNHLTLPDQWAHAGRLDVLAVAEPVTFSDMEVKITWGADTAACGDIISGLPVSARG 378

QY 182 REILGPADNFEGQWRLLAPITASQOTRGIGCITSITGRDKNOVEGEVQVSTATQ 241  
Db 379 KEILGPADSLERGWRLLAPITASQOTRGIGCITSITGRDKNOVEGEVQVSTATQ 438

QY 302 SDLYLVTRHADVTPVRGDSRSLSRSPVSYLKGSGGPLCPSGHAYGIFRAVCTR 361  
Db 499 SDLYLVTRHADVTPVRGDSRSLSRSPVSYLKGSGGPLCPSGHAYGIFRAVCTR 558

QY 362 GVAKAVDFPVESMETMR 380  
Db 559 GVAKAVDFPVESMETMR 577

RESULT 9  
US-09-919-901-2  
; Sequence 2, Application US/09263931  
; Patent No. 6599738  
; GENERAL INFORMATION:  
; APPLICANT: Potts, Karen E.  
; APPLICANT: Jackson, Roberta L.  
; APPLICANT: Potts, Karen E.  
; APPLICANT: Jackson, Roberta L.  
; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE  
; FILE REFERENCE: 0125-0005A  
; CURRENT APPLICATION NUMBER: US/09/919, 901  
; CURRENT FILING DATE: 2001-08-02  
; PRIOR APPLICATION NUMBER: 09/263, 933  
; PRIOR FILING DATE: 1999-02-08  
; PRIOR APPLICATION NUMBER: 09/129, 611  
; PRIOR FILING DATE: 1998-08-05  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 2307  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE: OTHER INFORMATION: :  
; US-09-919-901-2

Query Match 94.7%; Score 1881; DB 4; Length 2307;  
Best Local Similarity 93.1%; Pred. No. 1..7e-180; Mismatches 14; Indels 0; Gaps 0;  
Matches 353; Conservative 12; Mismatches 14; Indels 0; Gaps 0;

QY 2 LITLSPYYKVLLARLILWLOQITRVAHLQWIPPLNVRGRDAILITCAVHPELF 61  
Db 139 LITLSPYYKVFLARLILWLOQITRVAHLWIPPLNARGRDAILITCAVHPELF 258

QY 62 ITKLLAIFGPMVLTQAGITKPYFRAQGLIRACMVKAGGHVQMAFMKLAUTG 121  
Db 242 SPLATCNGVCTVYHGASKTKLAGPKPITOMYTNDQDVGWQAPPGRASMTCTCGS 301  
Db 439 SPLATCNGVCTVYHGASKTKLAGPKPITOMYTNDQDVGWQAPPGRASMTCTCGS 498

QY 122 YVDLHLPQLQWAHAGRLDVLAVAEPVTFSDMEVKITWGADTAACGDIISGLPVSARG 181  
Db 379 KEILGPADSLERGWRLLAPITASQOTRGIGCITSITGRDKNOVEGEVQVSTATQ 438

QY 242 SPLATCNGVCTVYHGASKTKLAGPKPITOMYTNDQDVGWQAPPGRASMTCTCGS 301  
Db 499 SDLYLVTRHADVTPVRGDSRSLSRSPVSYLKGSGGPLCPSGHAYGIFRAVCTR 558

QY 362 GVAKAVDFPVESMETMR 380  
Db 559 GVAKAVDFPVESMETMR 577

RESULT 10  
US-09-919-901-2  
; Sequence 2, Application US/09263931  
; Patent No. 6599738  
; GENERAL INFORMATION:  
; APPLICANT: Potts, Karen E.  
; APPLICANT: Jackson, Roberta L.  
; APPLICANT: Potts, Karen E.  
; APPLICANT: Jackson, Roberta L.  
; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE  
; FILE REFERENCE: 0125-0005A  
; CURRENT APPLICATION NUMBER: US/09/919, 901  
; CURRENT FILING DATE: 2001-08-02  
; PRIOR APPLICATION NUMBER: 09/263, 933  
; PRIOR FILING DATE: 1999-02-08  
; PRIOR APPLICATION NUMBER: 09/129, 611  
; PRIOR FILING DATE: 1998-08-05  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 2307  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE: OTHER INFORMATION: :  
; US-09-919-901-2

Query Match 94.7%; Score 1881; DB 4; Length 2307;  
Best Local Similarity 93.1%; Pred. No. 1..7e-180; Mismatches 14; Indels 0; Gaps 0;  
Matches 353; Conservative 12; Mismatches 14; Indels 0; Gaps 0;

QY 2 LITLSPYYKVLLARLILWLOQITRVAHLQWIPPLNVRGRDAILITCAVHPELF 61  
Db 139 LITLSPYYKVFLARLILWLOQITRVAHLWIPPLNARGRDAILITCAVHPELF 258

QY 62 ITKLLAIFGPMVLTQAGITKPYFRAQGLIRACMVKAGGHVQMAFMKLAUTG 121  
Db 242 SPLATCNGVCTVYHGASKTKLAGPKPITOMYTNDQDVGWQAPPGRASMTCTCGS 301  
Db 439 SPLATCNGVCTVYHGASKTKLAGPKPITOMYTNDQDVGWQAPPGRASMTCTCGS 498

QY 122 YVDLHLPQLQWAHAGRLDVLAVAEPVTFSDMEVKITWGADTAACGDIISGLPVSARG 181  
Db 379 KEILGPADSLERGWRLLAPITASQOTRGIGCITSITGRDKNOVEGEVQVSTATQ 438

QY 242 SPLATCNGVCTVYHGASKTKLAGPKPITOMYTNDQDVGWQAPPGRASMTCTCGS 301  
Db 499 SDLYLVTRHADVTPVRGDSRSLSRSPVSYLKGSGGPLCPSGHAYGIFRAVCTR 558

QY 362 GVAKAVDFPVESMETMR 380  
Db 559 GVAKAVDFPVESMETMR 577

RESULT 10  
US-09-539-601-33  
; Sequence 33, Application US/09539601C  
; Patent No. 6630343  
; GENERAL INFORMATION:  
; APPLICANT: Bartenschlager, Ralf FW  
; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System  
; FILE REFERENCE: all sequences  
; CURRENT APPLICATION NUMBER: US/09/539,601C  
; CURRENT FILING DATE: 2001-08-30  
; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY  
; EARLIER FILING DATE: 1995-04-03  
; NUMBER OF SEQ ID NOS: 51  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO: 33  
; LENGTH: 3010  
; TYPE: PRT  
; ORGANISM: Hepatitis C virus  
; US-09-539-601-33

Query Match 94.6%; Score 1880; DB 4; Length 3010;  
Best Local Similarity 93.1%; Pct. No. 3.1e-180; Mismatches 13; Indels 0; Gaps 0;  
Matches 353; Conservative 13; Mismatches 13; Indels 0; Gaps 0;

Qy 2 LTLSPSYYKVLARLWILQVYITRVAHLQWIPPLNVGGDRDAILTCAVHPBLIFD 61  
Db 828 LTLSPSYYKVLARLWILQVYITRVAHLQWIPPLNVGGDRDAILTCAVHPBLIFD 61  
Qy 62 ITKULLAIFGPMVNLQAGITKPYFVRAGLIRACMLVKRAAGHVVQAMKLAALTGT 121  
Db 888 ITKULLAIFGPMVNLQAGITKPYFVRAGLIRACMLVKRAAGHVVQAMKLAALTGT 947  
Qy 122 YVDHDLTPLQDWAFAGERDLAVAVEPVIFSDNEKVITWGAIDTAACGDISGLPVSARRG 181  
Db 948 YVDHDLTPLQDWAFAGERDLAVAVEPVIFSDNEKVITWGAIDTAACGDISGLPVSARRG 1007

Qy 182 REILLGPAINFEGCGWRLLAPITAYSOORTGIGCITSLGRDKQVEGEVQVSTATQ 241  
Db 287 KEILGPADSLIEGRGWRLAPITAYSOORTGIGCITSLGRDKQVEGEVQVSTATQ 346  
Qy 242 SFLATCVNGCWCWTYHGAGSKTLAGPKGPIITQMTNVDLWQAPPGRSMTPCTCGS 301  
Db 407 SDLYLVTRHADVPRRGDSRSLSPPRVSYLKGSGGPUCPSGHAVGIFRAAVCTR 361  
Qy 362 GVAKAVDFVYESMETMR 380  
Db 467 GVAKAVDFVYESMETMR 485

RESULT 11  
US-09-263-933-11  
; Sequence 11, Application US/09263933  
; Patent No. 6280340  
; GENERAL INFORMATION:  
; APPLICANT: Potts, Karen E.  
; APPLICANT: Jackson, Roberta L.  
; APPLICANT: Patrick, Amy K.  
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
; FILE REFERENCE: 0125-005A  
; CURRENT APPLICATION NUMBER: US/09/919,901  
; CURRENT FILING DATE: 2001-08-02  
; PRIORITY APPLICATION NUMBER: 09/263,933  
; PRIORITY FILING DATE: 1999-02-08  
; PRIORITY APPLICATION NUMBER: 09/129,611  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO: 11  
; LENGTH: 1692  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; OTHER INFORMATION: :  
; US-09-919-901-11

Query Match 94.5%; Score 1878; DB 4; Length 1692;

Best local Similarity 92.9%; Pred. No. 2.1e-180; Matches 352; Conservative 13; Mismatches 14; Indels 0; Gaps 0;

Db 379 KELLGPADSEGRHGRLLAPITASQFRGLCCTISLGRDKNOVEGEVQVSTATQ 438  
 Qy 2 LITLSPYKVFLARLIWMLQYLITRVEAHLOWIPNVRGGRDAILTCAVHPLEFD 61  
 Db 107 LITLSPYKVFLARLIWMLQYTTRABAHLYWIPNVRGGRDAILTCAVHPLEFD 166  
 Qy 62 ITKULLATFGPMLQAGITKPYFVRAGLIRACMLVKAGGHYVOMAFMKLAHTGT 121  
 Db 167 ITKULLATIGPLMVLQAGITKPYFVRAGLIRACMLVKAGGHYVOMAFMKLAHTGT 226  
 Qy 122 YVDHLPLQDWAHAGLDLAVAPEPVIFSDMEVKITTWGADTAACGDIISGLPVSARG 181  
 Db 227 YVNHLTPRLDWAHAGLDLAVAPEPVIFSDMEVKITTWGADTAACGDIILGPLVSARG 286  
 Qy 182 REILGPADNFGQWLAPITASQFRGLCCTISLGRDKNOVEGEVQVSTATQ 241  
 Db 287 KEILGPADSEGRGWRLLAPITASQFRGLCCTISLGRDKNOVEGEVQVSTATQ 346  
 Qy 242 SFLATCNGVCWTFHAGSKTLAGPKPQTOMTNVDLGWQAPPGRSMTCGS 301  
 Db 347 SFLATCNGVCWTFHAGSKTLAGPKPQTOMTNVDLGWQAPPGRSLPTCGS 406  
 Qy 302 SDLYLVTRHADVIPVRGRGDSRSLLSPRVSYLKGSAGGLCPSPGHAVGIFRAVCTR 361  
 Db 407 SDLYLVTRHADVIPVRGRGDSRSLLSPRVSYLKGSAGGLCPSPGHAVGIFRAVCTR 466  
 Qy 362 GVAKAVDFIPVESMETMR 380  
 Db 467 GVAKAVDFIPVESMETMR 485

RESULT 13  
 US-09-263-933-9  
 ; Sequence 9, Application US/09263933  
 ; Patent No. 628940  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Potts, Karen E.  
 ; APPLICANT: Jackson, Roberta L.  
 ; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
 ; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE  
 ; FILE REFERENCE: 0125-0005A  
 ; CURRENT APPLICATION NUMBER: US/09/919,901  
 ; CURRENT FILING DATE: 2001-08-02  
 ; PRIOR APPLICATION NUMBER: 09/263,933  
 ; PRIOR FILING DATE: 1999-02-08  
 ; PRIOR APPLICATION NUMBER: 09/129,611  
 ; PRIOR FILING DATE: 1998-08-05  
 ; NUMBER OF SEQ ID NOS: 33  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 9  
 ; LENGTH: 2307  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: :  
 ; US-09-919-901-9

Query Match 94.5%; Score 1878; DB 4; Length 2307;  
 Best Local Similarity 92.9%; Pred. No. 3.3e-180; Matches 352; Conservative 13; Mismatches 14; Indels 0; Gaps 0;

Db 379 KELLGPADSEGRHGRLLAPITASQFRGLCCTISLGRDKNOVEGEVQVSTATQ 438  
 Qy 2 LITLSPYKVFLARLIWMLQYLITRVEAHLOWIPNVRGGRDAILTCAVHPLEFD 61  
 Db 199 LITLSPYKVFLARLIWMLQYTTRABAHLYWIPNVRGGRDAILTCAVHPLEFD 258  
 Qy 62 ITKULLATFGPMLQAGITKPYFVRAGLIRACMLVKAGGHYVOMAFMKLAHTGT 121  
 Db 259 ITKULLATIGPLMVLQAGITKPYFVRAGLIRACMLVKAGGHYVOMAFMKLAHTGT 318  
 Qy 122 YVDHLPLQDWAHAGLDLAVAPEPVIFSDMEVKITTWGADTAACGDIISGLPVSARG 181  
 Db 319 YVNHLTPRLDWAHAGLDLAVAPEPVIFSDMEVKITTWGADTAACGDIILGPLVSARG 378  
 Qy 182 REILGPADNFGQWLAPITASQFRGLCCTISLGRDKNOVEGEVQVSTATQ 241  
 Db 349 SFLATCNGVCWTFHAGSKTLAGPKPQTOMTNVDLGWQAPPGRSLPTCGS 498  
 Db 379 KEILGPADSEGRGWRLLAPITASQFRGLCCTISLGRDKNOVEGEVQVSTATQ 438  
 Qy 242 SFLATCNGVCWTFHAGSKTLAGPKPQTOMTNVDLGWQAPPGRSMTCGS 301  
 Db 499 SDLYLVTRHADVIPVRGRGDSRSLLSPRVSYLKGSAGGLCPSPGHAVGIFRAVCTR 558  
 Qy 302 SDLYLVTRHADVIPVRGRGDSRSLLSPRVSYLKGSAGGLCPSPGHAVGIFRAVCTR 361  
 Db 499 SDLYLVTRHADVIPVRGRGDSRSLLSPRVSYLKGSAGGLCPSPGHAVGIFRAVCTR 558  
 Qy 362 GVAKAVDFIPVESMETMR 380  
 Db 559 GVAKAVDFIPVESMETMR 577

RESULT 14  
 US-09-919-901-9  
 ; Sequence 9, Application US/09919901  
 ; Patent No. 659978  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Potts, Karen E.  
 ; APPLICANT: Jackson, Roberta L.  
 ; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
 ; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE  
 ; FILE REFERENCE: 0125-0005A  
 ; CURRENT APPLICATION NUMBER: US/09/919,901  
 ; CURRENT FILING DATE: 2001-08-02  
 ; PRIOR APPLICATION NUMBER: 09/263,933  
 ; PRIOR FILING DATE: 1999-02-08  
 ; PRIOR APPLICATION NUMBER: 09/129,611  
 ; PRIOR FILING DATE: 1998-08-05  
 ; NUMBER OF SEQ ID NOS: 33  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 9  
 ; LENGTH: 2307  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: :  
 ; US-09-919-901-9

Query Match 94.5%; Score 1878; DB 4; Length 2307;  
 Best Local Similarity 92.9%; Pred. No. 3.3e-180; Matches 352; Conservative 13; Mismatches 14; Indels 0; Gaps 0;

Db 379 KELLGPADSEGRHGRLLAPITASQFRGLCCTISLGRDKNOVEGEVQVSTATQ 438  
 Qy 2 LITLSPYKVFLARLIWMLQYLITRVEAHLOWIPNVRGGRDAILTCAVHPLEFD 61  
 Db 199 LITLSPYKVFLARLIWMLQYTTRABAHLYWIPNVRGGRDAILTCAVHPLEFD 258  
 Qy 62 ITKULLATFGPMLQAGITKPYFVRAGLIRACMLVKAGGHYVOMAFMKLAHTGT 121  
 Db 259 ITKULLATIGPLMVLQAGITKPYFVRAGLIRACMLVKAGGHYVOMAFMKLAHTGT 318  
 Qy 122 YVDHLPLQDWAHAGLDLAVAPEPVIFSDMEVKITTWGADTAACGDIISGLPVSARG 181  
 Db 319 YVNHLTPRLDWAHAGLDLAVAPEPVIFSDMEVKITTWGADTAACGDIILGPLVSARG 378  
 Qy 182 REILGPADNFGQWLAPITASQFRGLCCTISLGRDKNOVEGEVQVSTATQ 241  
 Db 439 SFLATCNGVCWTFHAGSKTLAGPKPQTOMTNVDLGWQAPPGRSLPTCGS 498  
 Qy 302 SDLYLVTRHADVIPVRGRGDSRSLLSPRVSYLKGSAGGLCPSPGHAVGIFRAVCTR 361  
 Db 499 SDLYLVTRHADVIPVRGRGDSRSLLSPRVSYLKGSAGGLCPSPGHAVGIFRAVCTR 558  
 Qy 362 GVAKAVDFIPVESMETMR 380  
 Db 559 GVAKAVDFIPVESMETMR 577

RESULT 15  
US-09-263-933-18  
; Sequence 18, Application US/09263933

APPLICANT: Potts, Karen E.  
APPLICANT: Jackson, Roberta L.  
APPLICANT: Patick, Amy K.  
TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE  
FILE REFERENCE: 0125-0005A

EARLIER APPLICATION NUMBER: 09/129,611  
EARLIER FILING DATE: 1998-08-05  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO: 18  
166

-----  
; ORGANISM: Artificial Sequence  
; US-09-263-933-18

|         |   |       |              |           |              |
|---------|---|-------|--------------|-----------|--------------|
| Qy      | Query Match   | 94.1% | Score 1.689; | DB 3;     | Length 1692; |
| Matches | Best Local Similarity   | 92.6% | Pred No.     | 1.7e-179; |              |
| 351;    | Conservative  | 13;   | Mismatches   | 15;       | Indices 0;   |
| Db      |   |       | Gaps         | 0;        |              |
| Qy      | 2 LILTLSPSYKYLARLILWNLQYLITRVEAHLWQWLPPLNTRGRGRIAIIILTCAVHRELID   | 61    |              |           |              |
| Db      | 107 LILTLSPSYKYLARLILWNLQYLITRVEAHLWQWLPPLNTRGRGRIAIIILTCAVHRELID | 166   |              |           |              |
| Qy      | 62 ITKULLATIIGPMLYLAQAGITKVPYFVRAGLIRACMVLKAAGGYVQAFMKLAALTGT     | 121   |              |           |              |
| Db      | 167 ITKULLATIIGPMLYLAQAGITKVPYFVRAGLIRACMVLKAAGGYVQAFMKLAALTGT    | 226   |              |           |              |
| Qy      | 122 IYUDHALFLQDWAHAGJRLAVAEFPIVSDEMEKLITNGADTAAGDLSGLPSARRG       | 181   |              |           |              |
| Db      | 227 YIVNHHTLFIRDWAHAGJRLAVAEFPIVSDEMETKLITNGADTAAGDLSGLPSARRG     | 286   |              |           |              |
| Qy      | 182 RETILLGPADNFEGGWRLLAPITAYQSQTFRGLLGGIITSLTGRDQNQVEGEVQSTATO   | 241   |              |           |              |
| Db      | 287 KETILLGPADLSERGRWLAPITAYQSQTFRGLLGGIITSLTGRDQNQVEGEVQSTATO    | 346   |              |           |              |
| Qy      | 242 SFLATCNGVQCVWTFHAGSKTLAGPKPTQMYVNVDODLVGWAAPPGRASMRMPCTCGS    | 301   |              |           |              |
| Db      | 347 SFLATCNGVQCVWTFHAGSKTLAGPKPTQMYVNVDODLVGWAAPPGRASLRSLTCGS     | 406   |              |           |              |
| Qy      | 302 SDLYINTRHADIVFRRRGDSRGSLISPRSVYIKQSSGAPLICSGHAYGIIFRAVCTR     | 361   |              |           |              |
| Db      | 362 GVAKAVDFPVVESHEBTMR   | 380   |              |           |              |
| Qy      | 407 SDLYINTRHADIVFRRRGDSRGSLISPRSVYIKQSSGAPLICSGHAYGIIFRAVCTR     | 466   |              |           |              |
| Db      | 467 GVAKAVDFPVVESHEBTMR   | 485   |              |           |              |

Search completed: May 6, 2004, 09:39:03  
Job time : 15.9144 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on:

May 6, 2004, 09:05:56 ; Search time 38.8727 Seconds

(without alignments)  
2713.357 Million cell updates/sec

Title: score: US-10-650-585-12

Perfect score: 1987  
Sequence: ALITLSPYKYLARLIWL.....RGVAKAVDFIPVESMETTMR 380

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1140673 seqs, 277566755 residues

Total number of hits satisfying chosen parameters: 1140673

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/prodata/1/pupaa/us07\_PUBCOMB.pep:\*
- 2: /cgn2\_6/ptodata/1/pupaa/PCT\_NW\_PUB.pep:\*
- 3: /cgn2\_6/ptodata/1/pupaa/us06\_PUBCOMB.pep:\*
- 4: /cgn2\_6/ptodata/1/pupaa/us07\_NW\_PUB.pep:\*
- 5: /cgn2\_6/ptodata/1/pupaa/us07\_NW\_PUB.pep:\*
- 6: /cgn2\_6/ptodata/1/pupaa/PCTUS\_PUBCOMB.pep:\*
- 7: /cgn2\_6/ptodata/1/pupaa/us08\_NW\_PUB.pep:\*
- 8: /cgn2\_6/ptodata/1/pupaa/us08\_PUBCOMB.pep:\*
- 9: /cgn2\_6/ptodata/1/pupaa/us09\_PUBCOMB.pep:\*
- 10: /cgn2\_6/ptodata/1/pupaa/us09C\_PUBCOMB.pep:\*
- 11: /cgn2\_6/ptodata/1/pupaa/us09\_NW\_PUB.pep:\*
- 12: /cgn2\_6/ptodata/1/pupaa/us09\_PUBCOMB.pep:\*
- 13: /cgn2\_6/ptodata/1/pupaa/us10\_PUBCOMB.pep:\*
- 14: /cgn2\_6/ptodata/1/pupaa/us10\_PUBCOMB.pep:\*
- 15: /cgn2\_6/ptodata/1/pupaa/us10c\_PUBCOMB.pep:\*
- 16: /cgn2\_6/ptodata/1/pupaa/us10\_NW\_PUB.pep:\*
- 17: /cgn2\_6/ptodata/1/pupaa/us60\_NW\_PUB.pep:\*
- 18: /cgn2\_6/ptodata/1/pupaa/us60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result Score % Match Length DB ID Description

No. Score Query Match Length DB ID Description

Score Query Match Length DB ID Description

Match Score Query Match Length DB ID Description

Length Score Query Match Length DB ID Description

DB Score Query Match Length DB ID Description

ID Score Query Match Length DB ID Description

Description Score Query Match Length DB ID

DB Description Score Query Match Length DB ID

ID Description Score Query Match Length DB ID

Description DB Score Query Match Length DB ID

DB Description DB Score Query Match Length DB ID

ID Description DB Score Query Match Length DB ID

Description DB Description DB Score Query Match Length DB ID

DB Description DB Description DB Score Query Match Length DB ID

ID Description DB Description DB Score Query Match Length DB ID

Description DB Description DB Description DB Score Query Match Length DB ID

DB Description DB Description DB Description DB Score Query Match Length DB ID

ID Description DB Description DB Description DB Score Query Match Length DB ID

Description DB Description DB Description DB Description DB Score Query Match Length DB ID

DB Description DB Description DB Description DB Description DB Score Query Match Length DB ID

### RESULT 1

US-10-017-736-12

; Sequence 12, Application US/10017736

; Publication No. US20020192640A1

; GENERAL INFORMATION:

; APPLICANT: Boehringer Ingelheim (Canada) Ltd.

; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease

; FILE REFERENCE: 1/082

; CURRENT APPLICATION NUMBER: US10/017,736

; CURRENT FILING DATE: 2001-12-14

; PRIOR APPLICATION NUMBER: 60/256,031

; PRIOR FILING DATE: 2000-12-15

; NUMBER OF SEQ ID NOS: 21

; SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO: 12

; LENGTH: 380

; TYPE: PRT

; ORGANISM: HCV

; US-10-017-736-12

Query Match 100.0%; Score 1987; DB 13; Length 380;  
Best Local Similarity 100.0%; Pred. No. 1.2e-193; Length 380;  
Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

|    |  |
|----|--|
| Qy | 1 ALITLSPYKYLARLIWLQLYLITRVEHLQWIPPLNVRGRDAILLTCAVHPELF 60     |
|    |  |
| Db | 1 ALITLSPYKYLARLIWLQLYLITRVEHLQWIPPLNVRGRDAILLTCAVHPELF 60     |
| Ov | 61 DITKLILAFGLMVHQAGIKTKVPPFVRAQCLIRACMUTKAAGHYVQAFMKLAATG 120 |
|    |  |
| Db | 61 DITKLILAFGLMVHQAGIKTKVPPFVRAQCLIRACMUTKAAGHYVQAFMKLAATG 120 |
| Ov | 121 TTYDHDPLQWAHGIRDALAVAEPVIFSDMEKLTWGAACGDIISGLPSARR 180     |
| Db | 121 TTYDHDPLQWAHGIRDALAVAEPVIFSDMEKLTWGAACGDIISGLPSARR 180     |
| Ov | 181 GRELLGPADNEEGQSMRLPAPITAYSQORGULGCTTSITGRDNQZGEVQVSTA 240  |

```

RESULT 2
US-10-650-585-12
; Sequence 12, Application US/10650585
; Publication No. US20040077066A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
; CURRENT APPLICATION NUMBER: US/10/017,736
; CURRENT FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256,031
; PRIOR FILING DATE: 2000-12-15
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO: 11
; LENGTH: 393
; TYPE: PRT
; ORGANISM: HCV
; US-10-017-736-11

Query Match          100.0%; Score 1987; DB 16; Length 393;
Best Local Similarity 100.0%; Pred. No. 1; 3e-193; Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ALITISPYKVKLLARLWQLITRQYLTRVEARLQWNPPLAVRGGSDAIILITCAVHPBLIF 60
Db 14 ALITISPYKVKLLARLWQLITRQYLTRBAHLQWNPPLAVRGGRDAIILITCAVHPBLIF 73
QY 61 DITKULLAIFGPMVLPQKGLITKPYFVRAQGLRACMLVRKAGGHYQMAFKLALTG 120
Db 74 DITKULLAIFGPMVLPQKGLITKPYFVRAQGLRACMLVRKAGGHYQMAFKLALTG 133
QY 121 TYTYDHITPLQWMAHGLDIAVAVEVIVFSMEVKILTGWADTAACGDISGLPSARR 180
Db 134 TYTYDHITPLQWMAHGLDIAVAVEVIVFSMEVKILTGWADTAACGDISGLPSARR 193
QY 181 GREILLGPADNEFGQGLRPLAPTAQSQTQRLGLCITSIGRDKNQVEGEQVSTAT 240
Db 194 GREILLGPADNEFGQGLRPLAPTAQSQTQRLGLCITSIGRDKNQVEGEQVSTAT 253
QY 241 QSLFLATCNGVWTWFGAGSKTLAGPKGPITQMYNTDQLGWQAPPGRASMTPTCTG 300
Db 254 QSLFLATCNGVWTWFGAGSKTLAGPKGPITQMYNTDQLGWQAPPGRASMTPTCTG 313
QY 301 SSDLYLVTTHADVIPRRGDSRGSLSSPRVSYLKSSGGPLCPSGHAVGIFRAVCT 360
Db 314 SSDLYLVTTHADVIPRRGDSRGSLSSPRVSYLKSSGGPLCPSGHAVGIFRAVCT 373
Qy 361 RGVAKAVDFIPVESMETMR 380
Db 374 RGVAKAVDFIPVESMETMR 393

RESULT 4
US-10-650-585-11
; Sequence 11, Application US/10650585
; Publication No. US20040077066A1
; GENERAL INFORMATION:
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease
; FILE REFERENCE: 13/082
; CURRENT APPLICATION NUMBER: US/10/017,736
; CURRENT FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: 60/256,031
; PRIOR FILING DATE: 2000-12-15
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO: 11
; LENGTH: 393
; TYPE: PRT
; ORGANISM: HCV
; US-10-650-585-11

Query Match          100.0%; Score 1987; DB 16; Length 393;

```

Best local Similarity 100.0%; Pred. No. 1.3e-193; Mismatches 0; Indels 0; Gaps 0; Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ALLTISPYKVLLARLILWLOLITRVAHLQWIPPLNVRGRDAAITLTCAVHPELIF 60  
Db 14 ALLTISPYKVLLARLILWLOLITRVAHLQWIPPLNVRGRDAAITLTCAVHPELIF 73

QY 61 DITKULLAIFGPMLVQAGITKPYFVRAQIIRACMLVRZAGGYQMAFMKAALTG 120  
Db 74 DITKULLAIFGPMLVQAGITKPYFVRAQIIRACMLVRZAGGYQMAFMKAALTG 133

QY 121 TYVDHPLPQDWAHAGRLDAVAPVFSMEVKITWGDTAACGDIISGLPSARR 180  
Db 134 TYVDHPLPQDWAHAGRLDAVAPVFSMEVKITWGDTAACGDIISGLPSARR 193

QY 181 GRETLIGPADNFEGQWRILLAPITAYSQQRGLCITSUTGRDKNOQEVEQVSTAT 240  
Db 194 GRETLIGPADNFEGQWRILLAPITAYSQQRGLCITSUTGRDKNOQEVEQVSTAT 253

QY 241 QSLFLATCNGVCWTVFHAGSKTLAGPKPITOQNVDQDINGWQAPPGRSMPTCTG 300  
Db 254 QSLFLATCNGVCWTVFHAGSKTLAGPKPITOQNVDQDINGWQAPPGRSMPTCTG 313

QY 301 SSDLYLVTRHADVIPIVRRGSRSILSPRVSYIKGSSGGPLCPSGHVGIFRAVCT 360  
Db 318 SSDLYLVTRHADVIPIVRRGSRSILSPRVSYIKGSSGGPLCPSGHVGIFRAVCT 377

QY 361 RGVAKAVDPPIPESMETMR 380  
Db 378 RGVAKAVDPPIPESMETMR 397

QY 314 SSDLYLVTRHADVIPIVRRGSRSILSPRVSYIKGSSGGPLCPSGHVGIFRAVCT 373

QY 361 RGVAKAVDPPIPESMETMR 380  
Db 374 RGVAKAVDPPIPESMETMR 393

RESULT 5  
US-10-017-736-2  
; Sequence 2, Application US/10017736  
; Publication No. US2002012640A1  
; GENERAL INFORMATION:  
; APPLICANT: Boehringer Ingelheim (Canada) Ltd. Protease  
; TITLE OF INVENTION: Purified Active Hcv NS2/3 Protease  
; FILE REFERENCE: 13/082  
; CURRENT APPLICATION NUMBER: US/10/650,58  
; PRIORITY FILING DATE: 2003-08-28  
; PRIORITY APPLICATION NUMBER: US10/017,736A  
; PRIORITY FILING DATE: 2001-12-14  
; PRIORITY APPLICATION NUMBER: 60/256,031  
; PRIORITY FILING DATE: 2000-12-15  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 409  
; TYPE: PRY  
; ORGANISM: HCV  
; US-10-650-585-2

Query Match 100.0%; Score 1987; DB 13; Length 409;  
Best local Similarity 100.0%; Score 1987; DB 16; Length 409;  
Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ALLTISPYKVLLARLILWLOLITRVAHLQWIPPLNVRGRDAAITLTCAVHPELIF 60  
Db 18 ALLTISPYKVLLARLILWLOLITRVAHLQWIPPLNVRGRDAAITLTCAVHPELIF 77

QY 61 DITKULLAIFGPMLVQAGITKPYFVRAQIIRACMLVRZAGGYQMAFMKAALTG 120  
Db 78 DITKULLAIFGPMLVQAGITKPYFVRAQIIRACMLVRZAGGYQMAFMKAALTG 137

QY 121 TYVDHPLPQDWAHAGRLDAVAPVFSMEVKITWGDTAACGDIISGLPSARR 180  
Db 138 TYVDHPLPQDWAHAGRLDAVAPVFSMEVKITWGDTAACGDIISGLPSARR 197

QY 181 GRETLIGPADNFEGQWRILLAPITAYSQQRGLCITSUTGRDKNOQEVEQVSTAT 240  
Db 198 GRETLIGPADNFEGQWRILLAPITAYSQQRGLCITSUTGRDKNOQEVEQVSTAT 257

QY 241 QSLFLATCNGVCWTVFHAGSKTLAGPKPITOQNVDQDINGWQAPPGRSMPTCTG 300  
Db 258 QSLFLATCNGVCWTVFHAGSKTLAGPKPITOQNVDQDINGWQAPPGRSMPTCTG 317

QY 301 SSDLYLVTRHADVIPIVRRGSRSILSPRVSYIKGSSGGPLCPSGHVGIFRAVCT 360  
Db 318 SSDLYLVTRHADVIPIVRRGSRSILSPRVSYIKGSSGGPLCPSGHVGIFRAVCT 377

QY 361 RGVAKAVDPPIPESMETMR 380  
Db 378 RGVAKAVDPPIPESMETMR 397

RESULT 6  
US-10-650-585-2  
; Sequence 2, Application US/10650585  
; Publication No. US2004007706A1  
; GENERAL INFORMATION:  
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.  
; TITLE OF INVENTION: Purified Active Hcv NS2/3 Protease  
; FILE REFERENCE: 13/082  
; CURRENT APPLICATION NUMBER: US/10/650,58  
; PRIORITY FILING DATE: 2003-08-28  
; PRIORITY APPLICATION NUMBER: US10/017,736A  
; PRIORITY FILING DATE: 2001-12-14  
; PRIORITY APPLICATION NUMBER: 60/256,031  
; PRIORITY FILING DATE: 2000-12-15  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 409  
; TYPE: PRY  
; ORGANISM: HCV  
; US-10-650-585-2

Query Match 100.0%; Score 1987; DB 13; Length 409;  
Best local Similarity 100.0%; Score 1987; DB 16; Length 409;  
Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ALLTISPYKVLLARLILWLOLITRVAHLQWIPPLNVRGRDAAITLTCAVHPELIF 60  
Db 18 ALLTISPYKVLLARLILWLOLITRVAHLQWIPPLNVRGRDAAITLTCAVHPELIF 77

QY 61 DITKULLAIFGPMLVQAGITKPYFVRAQIIRACMLVRZAGGYQMAFMKAALTG 120  
Db 78 DITKULLAIFGPMLVQAGITKPYFVRAQIIRACMLVRZAGGYQMAFMKAALTG 137

QY 121 TYVDHPLPQDWAHAGRLDAVAPVFSMEVKITWGDTAACGDIISGLPSARR 180  
Db 138 TYVDHPLPQDWAHAGRLDAVAPVFSMEVKITWGDTAACGDIISGLPSARR 197

QY 181 GRETLIGPADNFEGQWRILLAPITAYSQQRGLCITSUTGRDKNOQEVEQVSTAT 240  
Db 198 GRETLIGPADNFEGQWRILLAPITAYSQQRGLCITSUTGRDKNOQEVEQVSTAT 257

QY 241 QSLFLATCNGVCWTVFHAGSKTLAGPKPITOQNVDQDINGWQAPPGRSMPTCTG 300  
Db 258 QSLFLATCNGVCWTVFHAGSKTLAGPKPITOQNVDQDINGWQAPPGRSMPTCTG 317

QY 301 SSDLYLVTRHADVIPIVRRGSRSILSPRVSYIKGSSGGPLCPSGHVGIFRAVCT 360  
Db 318 SSDLYLVTRHADVIPIVRRGSRSILSPRVSYIKGSSGGPLCPSGHVGIFRAVCT 377

QY 361 RGVAKAVDPPIPESMETMR 380  
Db 378 RGVAKAVDPPIPESMETMR 397

RESULT 7  
US-10-029-907-3  
; Sequence 3, Application US/10029907  
; Publication No. US20020142350A1  
; GENERAL INFORMATION:  
; APPLICANT: BOEHRINGER INGELHEIM (CANADA) LTD.



Matches 354; Conservative 13; Mismatches 12; Indels 0; Gaps 0;

Qy 2 LITLSPYKVLLARLILWQLQTIRVEAHQWIPNTRGGRAILITCAVHEPILFD 61  
Db 828 LITLSPHYKFLARLILWQLQTIRVEAHQWIPNTRGGRAILITCAVHEPILFT 837

Qy 62 ITKULLAIGPLMLQAGITKPVFVRQGLIRACMVLVRKAGGHVQMAFMKLAUTG 121  
Db 888 ITKULLAIGPLMLQAGITKPVFVRQGLIRACMVLVRKAGGHVQMAFMKLAUTG 947

Qy 122 YVDHITPLQDWAHAGRLDAVAVERVIFSDMEVKITWGAACCDISGGPSARRG 181  
Db 948 YVDHITPLQDWAHAGRLDAVAVERVIFSDMEVKITWGAACCDISGGPSARRG 1007

Qy 182 REILIGPADNFGQGRLLAPITASQOTGLGCITSQTRGLGCITSLGRDRNQVEGVQVSTATQ 241  
Db 1008 REILIGPADNFGQGRLLAPITASQOTGLGCITSQTRGLGCITSLGRDRNQVEGVQVSTATQ 1067

Qy 242 SPLATCNGVWTFHAGSKTLAGPKPTOMYNTWDQDLYGWQAPPGRASMPCTCGS 301  
Db 1068 SPLATCNGVWTFHAGSKTLAGPKPTOMYNTWDQDLYGWQAPPGRASMPCTCGS 1127

Qy 302 SDLYLVTRHADVIPVRRGDSRSGLSPRPSYIKGSSGGPLCPSGHAVGIFRAAVCTR 361  
Db 1128 SDLYLVTRHADVIPVRRGDSRSGLSPRPSYIKGSSGGPLCPSGHAVGIFRAAVCTR 1187

Qy 362 GVAKAVDFPVESMETMR 380  
Db 1188 GVAKAVDFPVESMETMR 1206

RESULT 10  
US-09-919\_901-4  
Sequence 4, Application US/09919901  
Publication No. US20030082518A1  
GENERAL INFORMATION:  
APPLICANT: Potts, Karen E.  
APPLICANT: Jackson, Roberta L.  
APPLICANT: Patrick, Amy K.  
TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
FILE REFERENCE: 0125-0005A  
CURRENT APPLICATION NUMBER: US/10/191\_966  
CURRENT FILING DATE: 2002-07-10  
PRIOR APPLICATION NUMBER: US/09/263, 933  
PRIOR FILING DATE: 1999-03-08  
PRIOR APPLICATION NUMBER: 09/129, 611  
PRIOR FILING DATE: 1998-08-05  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 4  
LENGTH: 1692  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION:  
US-10-191-966-4

Query Match 94.7%; Score 1881; DB 14; Length 1692;  
Best Local Similarity 93.1%; Pred. No. 6.4e-182; Mismatches 12; Indels 0; Gaps 0;

Matches 353; Conservative 12; Mismatches 14; Indels 0; Gaps 0;

Qy 2 LITLSPYKVLLARLILWQLQTIRVEAHQWIPNTRGGRAILITCAVHEPILFD 61  
Db 107 LITLSPYKVFLARLILWQLQTIRVEAHQWIPNTRGGRAILITCAVHEPILFD 166

Qy 62 ITKULLAIGPLMLQAGITKPVFVRQGLIRACMVLVRKAGGHVQMAFMKLAUTG 121  
Db 167 ITKULLAIGPLMLQAGITKPVFVRQGLIRACMVLVRKAGGHVQMAFMKLAUTG 226

Qy 122 YVDHITPLQDWAHAGRLDAVAVERVIFSDMEVKITWGAACCDISGGPSARRG 181  
Db 227 YVDHITPLQDWAHAGRLDAVAVERVIFSDMEVKITWGAACCDISGGPSARRG 286

Qy 242 SPLATCNGVWTFHAGSKTLAGPKPTOMYNTWDQDLYGWQAPPGRASMPCTCGS 301  
Db 347 SPLATCNGVWTFHAGSKTLAGPKPTOMYNTWDQDLYGWQAPPGRASMPCTCGS 406

Qy 302 SDLYLVTRHADVIPVRRGDSRSGLSPRPSYIKGSSGGPLCPSGHAVGIFRAAVCTR 361  
Db 407 SDLYLVTRHADVIPVRRGDSRSGLSPRPSYIKGSSGGPLCPSGHAVGIFRAAVCTR 466

Qy 362 GVAKAVDFPVESMETMR 380  
Db 467 GVAKAVDFPVESMETMR 485

RESULT 11  
US-10-191-966-4  
Sequence 4, Application US/10/191966  
Publication No. US20030175692A1  
GENERAL INFORMATION:  
APPLICANT: Potts, Karen E.  
APPLICANT: Jackson, Roberta L.  
APPLICANT: Patrick, Amy K.  
TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
FILE REFERENCE: 0125-0005A  
CURRENT APPLICATION NUMBER: US/10/191\_966  
CURRENT FILING DATE: 2002-07-10  
PRIOR APPLICATION NUMBER: US/09/263, 933  
PRIOR FILING DATE: 1999-03-08  
PRIOR APPLICATION NUMBER: 09/129, 611  
PRIOR FILING DATE: 1998-08-05  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 4  
LENGTH: 1692  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION:  
US-10-191-966-4

Query Match 94.7%; Score 1881; DB 14; Length 1692;  
Best Local Similarity 93.1%; Pred. No. 6.4e-182; Mismatches 12; Indels 0; Gaps 0;

Matches 353; Conservative 12; Mismatches 14; Indels 0; Gaps 0;

Qy 2 LITLSPYKVLLARLILWQLQTIRVEAHQWIPNTRGGRAILITCAVHEPILFD 61  
Db 107 LITLSPYKVFLARLILWQLQTIRVEAHQWIPNTRGGRAILITCAVHEPILFD 166

Qy 62 ITKULLAIGPLMLQAGITKPVFVRQGLIRACMVLVRKAGGHVQMAFMKLAUTG 121  
Db 167 ITKULLAIGPLMLQAGITKPVFVRQGLIRACMVLVRKAGGHVQMAFMKLAUTG 226

Qy 122 YVDHITPLQDWAHAGRLDAVAVERVIFSDMEVKITWGAACCDISGGPSARRG 181  
Db 227 YVDHITPLQDWAHAGRLDAVAVERVIFSDMEVKITWGAACCDISGGPSARRG 286

Qy 242 SPLATCNGVWTFHAGSKTLAGPKPTOMYNTWDQDLYGWQAPPGRASMPCTCGS 301  
Db 347 SPLATCNGVWTFHAGSKTLAGPKPTOMYNTWDQDLYGWQAPPGRASMPCTCGS 406

Qy 302 SDLYLVTRHADVIPVRRGDSRSGLSPRPSYIKGSSGGPLCPSGHAVGIFRAAVCTR 361  
Db 407 SDLYLVTRHADVIPVRRGDSRSGLSPRPSYIKGSSGGPLCPSGHAVGIFRAAVCTR 466

QY 362 GVAKAVDPFIPVSMETMR 380  
Db 467 GVAKAVDPFIPVSMETMR 485

RESULT 12  
US-09-919-901-2  
; Sequence 2, Application US/09919901  
; Publication No. US2003008251B1  
; GENERAL INFORMATION:  
; APPLICANT: Potts, Karen E.  
; APPLICANT: Patick, Amy K.  
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
FILE REFERENCE: 0125-0005A  
CURRENT APPLICATION NUMBER: US/09/919, 901  
CURRENT FILING DATE: 2001-08-02  
PRIORITY NUMBER: US/09/263, 933  
PRIORITY NUMBER: 09/129, 611  
PRIORITY FILING DATE: 1998-08-05  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO: 2  
LENGTH: 2307  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE: OTHER INFORMATION :  
OTHER INFORMATION :  
US-09-919-901-2

Query Match 94.7%; Score 1881; DB 10; Length 2307;  
Best Local Similarity 93.1%; Pred. No. 1e-181; Matches 353; Conservative 12; Mismatches 14; Indels 0; Gaps 0;  
Matches 353; Conservative 12; Mismatches 14; Indels 0; Gaps 0;

QY 2 LITLSPYKYLARLWILQQLYLTIVRAHQWIPPLNVRGRDAILTCAVHPSELFD 61  
Db 199 LITLSPYKYLARLWILQQLYLTIVRAHQWIPPLNVRGRDAILTCAVHPSELFD 258

QY 62 ITKULLALIFGPLMLQAGITRPFVRAQGLIACMLVRKAGHVVQAMAFKLAALTGT 121  
Db 259 ITKULLALIFGPLMLQAGITRPFVRAQGLIACMLVRKAGHVVQAMAFKLAALTGT 318

QY 122 YVDHDLTPIODWHAIGRDLLAVAPVPEPFSMDVKUTTWGADTAACGDIISGLPVARRG 181  
Db 319 YVDHDLTPIODWHAIGRDLLAVAPVPEPFSMDVKUTTWGADTAACGDIISGLPVARRG 378

QY 182 REILGPDPNPFEGQWRILAPITYSQTQTRGLGCTTSLGCRDKNOVEGEVQVSTATQ 241  
Db 379 KEILLGPADSLERGWRLAPIATAYSOOTRGCLGCTTSLGCRDKNOVEGEVQVSTATQ 438

QY 242 SFATCVCVQCVTFPHGAGSKTLAGPKPITQMYTNVDQDINGWQARPGASMTPTCGS 301  
Db 439 SFATCVCVQCVTFPHGAGSKTLAGPKPITQMYTNVDQDINGWQARPGASMTPTCGS 498

QY 302 SDLVLYTREADVPRRGDSRSLSLSPRPVSYLKQSSGGPLCPSGHAVGIFRAVCTR 361  
Db 499 SDLVLYTREADVPRRGDSRSLSLSPRPVSYLKQSSGGPLCPSGHAVGIFRAVCTR 558

QY 362 GVAKAVDPFIPVSMETMR 380  
Db 559 GVAKAVDPFIPVSMETMR 577

RESULT 14  
US-09-919-901-11  
; Sequence 11, Application US/09919901  
; Publication No. US2003008251B1  
; GENERAL INFORMATION:  
; APPLICANT: Potts, Karen E.  
; APPLICANT: Patick, Amy K.  
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
FILE REFERENCE: 0125-0005A  
CURRENT FILING DATE: 2001-08-02  
PRIORITY NUMBER: 09/263, 933  
PRIORITY NUMBER: US/09/263, 933  
PRIORITY NUMBER: 09/129, 611  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO: 11  
LENGTH: 1692

RESULT 13  
US-10-191-966-2  
; Sequence 2, Application US/10191966  
; Publication No. US20030175692A1  
; GENERAL INFORMATION:  
; APPLICANT: Potts, Karen E.  
; APPLICANT: Patick, Roberta L.  
; APPLICANT: Patick, Amy K.

```

; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE: OTHER INFORMATION: : US-09-919-901-11
; Query Match 94.5%; Score 1878; DB 10; Length 1692;
; Best Local Similarity 92.9%; Pred. No. 1 3e-181; Matches 352; Conservative 13; Mismatches 14; Indels 0; Gaps 0;
; Qy 2 LLLTSPYKYLARLIWNLQYLTTRVHLQWIPPINVRGGRDAILITLTCVHPLIFD 61
; Db 107 LLTSPYKYLARLIWNLQYLTTRVHLQWIPPINVRGGRDAILITLTCVHPLIFD 166
; Qy 62 ITKULLAIFGPLMLVQAGITKUPVFRAQGLIRACMLVKAAGCHYOMAFMKLAALTGT 121
; Db 167 ITKULLAIGPLMLVQAGITKUPVFRAQGLIRACMLVKAAGCHYOMAFMKLAALTGT 226
; Qy 122 YVDHMLPQDWAHAGLRLAVAVEPVNISDMEVKLTGADTAACGDTISGLPVSARG 181
; Db 227 YVNHLTPRDWAHAGLRLAVAVEPVNISDMEVKLTGADTAACGDTISGLPVSARG 286
; Qy 182 REILLGPADNFEGQWRLALAPITASQQTRGGLGCGTSITGRDNQVGEVQVSTATQ 241
; Db 167 ITKULLAIGPLMLVQAGITKUPVFRAQGLIRACMLVKAAGCHYOMAFMKLAALTGT 226
; Qy 122 YVDHMLPQDWAHAGLRLAVAVEPVNISDMEVKLTGADTAACGDTISGLPVSARG 181
; Db 227 YVNHLTPRDWAHAGLRLAVAVEPVNISDMEVKLTGADTAACGDTISGLPVSARG 286
; Qy 62 ITKULLAIFGPLMLVQAGITKUPVFRAQGLIRACMLVKAAGCHYOMAFMKLAALTGT 121
; Db 167 ITKULLAIGPLMLVQAGITKUPVFRAQGLIRACMLVKAAGCHYOMAFMKLAALTGT 226
; Qy 182 REILLGPADNFEGQWRLALAPITASQQTRGGLGCGTSITGRDNQVGEVQVSTATQ 241
; Db 167 ITKULLAIGPLMLVQAGITKUPVFRAQGLIRACMLVKAAGCHYOMAFMKLAALTGT 226
; Qy 242 SFLATCVNGCWTWGHGSKTLAGPKGITOMTNVDDLVGQAPPARSMTPTCGS 301
; Db 347 SFLATCVNGCWTWGHGSKTLAGPKGITOMTNVDDLVGQAPPARSMTPTCGS 406
; Qy 302 SDIVLYVTRHADVIPVRRGDSRSLSPRPVSYLKGSGGCPLCPSGHAGVIFRAVCTR 361
; Db 407 SDIVLYVTRHADVIPVRRGDSRSLSPRPVSYLKGSAGCPLCPSGHAGVIFRAVCTR 466
; Qy 362 GVAKAVDPVPESMETMR 380
; Db 467 GVAKAVDPVPESMETMR 485
; Search completed: May 6, 2004, 09:43:20
; Job time : 39.877 secs
; RESULT 15
; US-10-191-956-11
; Sequence 11, Application US/0191966
; Publication No. US20030175692A1
; GENERAL INFORMATION:
; APPLICANT: Potts, Karen E.
; APPLICANT: Jackson, Roberta L.
; APPLICANT: Patrick, Amy K.
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT
; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE
; FILE REFERENCE: 0125-0005A
; CURRENT APPLICATION NUMBER: US/10/191,966
; CURRENT FILING DATE: 2002-07-10
; PRIOR APPLICATION NUMBER: US/09/163,933
; PRIOR FILING DATE: 1999-03-08
; PRIOR APPLICATION NUMBER: 09/129,611
; PRIOR FILING DATE: 1998-08-05
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 1692
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE: OTHER INFORMATION: :
; US-10-191-956-11
; Query Match 94.5%; Score 1878; DB 14; Length 1692;
; Best Local Similarity 92.9%; Pred. No. 1 3e-181; Matches 352; Conservative 13; Mismatches 14; Indels 0; Gaps 0;
; Qy 2 LLLTSPYKYLARLIWNLQYLTTRVHLQWIPPINVRGGRDAILITLTCVHPLIFD 61
; Db 107 LLTSPYKYLARLIWNLQYLTTRVHLQWIPPINVRGGRDAILITLTCVHPLIFD 166

```



|  |   |       |              |       |             |   |           |                                     |   |   |   |
|--|---|-------|--------------|-------|-------------|---|-----------|-------------------------------------|---|---|---|
| Copyright (c) 1993 - 2004 Compugen Ltd.  | GenCore version 5.1.6                               |       |              |       |             |   |           |                                     |   |   |   |
| OM protein - protein search, using sw model  | (without alignments)                                |       |              |       |             |   |           |                                     |   |   |   |
| Run on:  | May 6, 2004, 09:08:45 ; Search time 47,9132 Seconds |       |              |       |             |   |           |                                     |   |   |   |
| Title:   | US-10-650-585-13                                    |       |              |       |             |   |           |                                     |   |   |   |
| Perfect score:   | 1842  |       |              |       |             |   |           |                                     |   |   |   |
| Sequence:  | AHLQWVWIPPLNVRGGRDAI.....RGAVAKAVDIFPVESMETM 352    |       |              |       |             |   |           |                                     |   |   |   |
| Scoring table:   | BLOSUM62  |       |              |       |             |   |           |                                     |   |   |   |
| Gapop:   | 10.0 , Gapext: 0.5                                  |       |              |       |             |   |           |                                     |   |   |   |
| Searched:  | 1586107 seqs, 28547505 residues                     |       |              |       |             |   |           |                                     |   |   |   |
| Total number of hits satisfying chosen parameters:   | 1586107   |       |              |       |             |   |           |                                     |   |   |   |
| Minimum DB seq length:   | 0   |       |              |       |             |   |           |                                     |   |   |   |
| Maximum DB seq length:   | 200000000   |       |              |       |             |   |           |                                     |   |   |   |
| Post-processing:   | Minimum Match 0*                                    |       |              |       |             |   |           |                                     |   |   |   |
|  | Maximum Match 100*                                  |       |              |       |             |   |           |                                     |   |   |   |
|  | Listing first 45 summaries                          |       |              |       |             |   |           |                                     |   |   |   |
| Database :   | A_Geneseq_29Jan04:*                                 |       |              |       |             |   |           |                                     |   |   |   |
| 1:   | geneseqp1980s:*                                     |       |              |       |             |   |           |                                     |   |   |   |
| 2:   | geneseqp1990s:*                                     |       |              |       |             |   |           |                                     |   |   |   |
| 3:   | geneseqp2000s:*                                     |       |              |       |             |   |           |                                     |   |   |   |
| 4:   | geneseqp2001s:*                                     |       |              |       |             |   |           |                                     |   |   |   |
| 5:   | geneseqp2002s:*                                     |       |              |       |             |   |           |                                     |   |   |   |
| 6:   | geneseqp2003as:*                                    |       |              |       |             |   |           |                                     |   |   |   |
| 7:   | geneseqp2003bs:*                                    |       |              |       |             |   |           |                                     |   |   |   |
| 8:   | geneseqp2004s:*                                     |       |              |       |             |   |           |                                     |   |   |   |
| pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution. |   |       |              |       |             |   |           |                                     |   |   |   |
| SUMMARIES  |   |       |              |       |             |   |           |                                     |   |   |   |
| Result No.   | Score   | Query | Match Length | DB ID | Description |   |           |                                     |   |   |   |
| 1.   | 1842  | 100.0 | 352          | 5     | ABG32186    | RESULT 1  | ABG32186  | ABG32186 standard; protein; 352 AA. | XX  | XX  |   |
| 2.   | 1842  | 100.0 | 380          | 5     | ABG32185    | ID ABG32186;  | ABG32186; | AC                                  | XX  | XX  |   |
| 3.   | 1842  | 100.0 | 393          | 5     | ABG32184    | DT 05-NOV-2002 (first entry)  | XX        | XX                                  | DE  | HCV protease NS2/3 truncation mutant 855-1206.  |   |
| 4.   | 1842  | 100.0 | 409          | 5     | ABG32181    | XX  | XX        | XX                                  | XX  | HCV; enzyme; protease; NS2/3 (855-1206); hepatitis C virus infection; chronic liver disease; end-stage liver disease; virucide; hepatotropic; antiinflammatory; lauryldiethylamine oxide; LD50; chaotropic agent; mutant; mutein. |   |
| 5.   | 1778  | 96.5  | 341          | 5     | ABG32187    | DE ABG32187 HCV prote   | ABG32187  | ABG32187                            | XX  | OS Hepatitis C virus.   |   |
| 6.   | 1773  | 96.3  | 3010         | 2     | ABG32694    | OS Synthetic.   | ABG32694  | ABG32694 Partial H                  | XX  | OS Hepatitis C virus.   |   |
| 7.   | 1768  | 96.0  | 3010         | 2     | ABG32622    | PD 20-JUN-2002.   | ABG32622  | ABG32622 HCV Prote                  | XX  | PD 20-JUN-2002.   |   |
| 8.   | 1767  | 95.9  | 3010         | 2     | ABG32864    | PF 13-DEC-2001; 2001WO-CA001796.  | ABG32864  | ABG32864 Hepatitis                  | XX  | PF 13-DEC-2001; 2001WO-CA001796.  |   |
| 9.   | 1766  | 95.9  | 2201         | 5     | ABG30601    | PR 15-DEC-2000; 2000US-0256031P.  | ABG30601  | ABG30601 Hepatitis                  | XX  | PR 15-DEC-2000; 2000US-0256031P.  |   |
| 10.  | 1766  | 95.9  | 2201         | 5     | ABG30591    | PA (BOEHINGER INGELHEIM CANADA LTD.   | ABG30591  | ABG30591 Hepatitis                  | XX  | PA (BOEHINGER INGELHEIM CANADA LTD.   |   |
| 11.  | 1766  | 95.9  | 2201         | 5     | ABG30500    | PI Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;  | ABG30500  | ABG30500 Hepatitis                  | XX  | PI Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;  |   |
| 12.  | 1766  | 95.9  | 2201         | 5     | ABG30581    | DR WPI; 2002-599511/64.   | ABG30581  | ABG30581 Hepatitis                  | XX  | DR WPI; 2002-599511/64.   |   |
| 13.  | 1766  | 95.9  | 2201         | 5     | ABG30593    | PS Claim 41; Page 61-62; 67pp; English.   | ABG30593  | ABG30593 Hepatitis                  | XX  | PS Claim 41; Page 61-62; 67pp; English.   |   |
| 14.  | 1766  | 95.9  | 2201         | 5     | ABG30582    | PT Novel polypeptide for screening inhibitors of non-structural proteases useful as therapeutic agents against hepatitis C virus, comprises full length non-structural protease, or its truncation.   | ABG30582  | ABG30582 Hepatitis                  | XX  | PT Novel polypeptide for screening inhibitors of non-structural proteases useful as therapeutic agents against hepatitis C virus, comprises full length non-structural protease, or its truncation.                               |   |
| 15.  | 1766  | 95.9  | 2201         | 5     | ABG30580    | PT Claim 41; Page 61-62; 67pp; English.   | ABG30580  | ABG30580 Hepatitis                  | XX  | PT Claim 41; Page 61-62; 67pp; English.   |   |
| 16.  | 1766  | 95.9  | 2201         | 5     | ABG30587    | CC The invention relates to an isolated polypeptide consisting of a full-length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal residue amino acid 810 to 906, or having a minimal amino acid sequence from residues 901 to 1206 of hepatitis C virus (HCV) 1b-40 full-length NS2/3 protease. Also included are (1) a composition (C) comprising an isolated HCV NS2/3 protease selected from full length NS2/3 protease, or its truncation, or a mutated sequence, where the protease is in a solution comprising a sufficient concentration of lauryldiethylamine oxide (LDAO) to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide | ABG30587  | ABG30587 Hepatitis                  | CC The invention relates to an isolated polypeptide consisting of a full-length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal residue amino acid 810 to 906, or having a minimal amino acid sequence from residues 901 to 1206 of hepatitis C virus (HCV) 1b-40 full-length NS2/3 protease. Also included are (1) a composition (C) comprising an isolated HCV NS2/3 protease selected from full length NS2/3 protease, or its truncation, or a mutated sequence, where the protease is in a solution comprising a sufficient concentration of lauryldiethylamine oxide (LDAO) to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide | XX  | CC The invention relates to an isolated polypeptide consisting of a full-length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal residue amino acid 810 to 906, or having a minimal amino acid sequence from residues 901 to 1206 of hepatitis C virus (HCV) 1b-40 full-length NS2/3 protease. Also included are (1) a composition (C) comprising an isolated HCV NS2/3 protease selected from full length NS2/3 protease, or its truncation, or a mutated sequence, where the protease is in a solution comprising a sufficient concentration of lauryldiethylamine oxide (LDAO) to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide |
| 17.  | 1766  | 95.9  | 2201         | 5     | ABG30594    | CC Abg32460 Hepatitis   | ABG32460  | Abg32460 Hepatitis                  | XX  | CC Abg32460 Hepatitis   |   |
| 18.  | 1766  | 95.9  | 2201         | 5     | ABG30598    | Abg32461 Hepatitis  | ABG32461  | Abg32461 Hepatitis                  | XX  | Abg32461 Hepatitis  |   |
| 19.  | 1766  | 95.9  | 2201         | 5     | ABG30599    | Abg32454 Hepatitis  | ABG32454  | Abg32454 Hepatitis                  | XX  | Abg32454 Hepatitis  |   |
| 20.  | 1766  | 95.9  | 2201         | 5     | ABG30595    | Abg32456 Hepatitis  | ABG32456  | Abg32456 Hepatitis                  | XX  | Abg32456 Hepatitis  |   |
| 21.  | 1766  | 95.9  | 3010         | 5     | ABG32458    | Abg3055 Hepatitis   | ABG3055   | Abg3055 Hepatitis                   | XX  | Abg3055 Hepatitis   |   |
| 22.  | 1766  | 95.9  | 3010         | 5     | ABG32459    | Abg3058 Hepatitis   | ABG3058   | Abg3058 Hepatitis                   | XX  | Abg3058 Hepatitis   |   |
| 23.  | 1766  | 95.9  | 3010         | 5     | ABG32451    | Abg3059 Hepatitis   | ABG3059   | Abg3059 Hepatitis                   | XX  | Abg3059 Hepatitis   |   |
| 24.  | 1766  | 95.9  | 3010         | 5     | ABG32455    | Abg32455 Hepatitis  | ABG32455  | Abg32455 Hepatitis                  | XX  | Abg32455 Hepatitis  |   |
| 25.  | 1766  | 95.9  | 3010         | 5     | ABG32457    | Abg32457 Hepatitis  | ABG32457  | Abg32457 Hepatitis                  | XX  | Abg32457 Hepatitis  |   |



|   |     |   |
|---|-----|---|
| RESULT 3  | Db  | QTRGLGCGITSLTGRDKNQEGEVQVSTATQSFLATCNGVWTFHGAGSKTLAGPK 268    |
| ID ABG32184   | 241 | GPIIOMTNVQDQLVWQAPGARSMPTCTCGSSDLYLVTTRHADIVFVRREDSRGSIIS 300 |
| XX ABG32184 standard; protein; 393 AA.  | 269 | GPIIOMTNVQDQLVWQAPGARSMPTCTCGSSDLYLVTTRHADIVFVRREDSRGSIIS 328 |
| AC ABG32184;  | 301 | PRPVSYIKGSQQGPILCPSGHAWGVIFRAAVCTRGVAKAVDFIPVESMETMR 352      |
| XX  | 329 | PRPVSYIKGSQQGPILCPSGHAWGVIFRAAVCTRGVAKAVDFIPVESMETMR 380      |
| DT 05-NOV-2002 (first entry)  |     |   |
| XX DE HCV protease NS2/3 truncation mutant 815-1206.  |     |   |
| XX HCV enzyme; protease; NS2/3 (815-1206); hepatitis C virus infection; chronic liver disease; cirrhosis; end-stage liver disease; virucide; hepatotropic; anti-inflammatory; lauryldimethylamine oxide; LDLO; chaotropic agent; mutant; mutein.  |     |   |
| XX KW Hepatitis C virus.  |     |   |
| OS Synthetic.   |     |   |
| XX PN WO200243375-A2.   |     |   |
| XX PD 20-JUN-2002.  |     |   |
| XX PF 13-DEC-2001; 2001WO-CA001796.   |     |   |
| XX PR 15-DEC-2000; 200001S-025031P.   |     |   |
| XX PA (BOHR ) BOEHRINGER INGELHEIM CANADA LTD.  |     |   |
| XX PI Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;   |     |   |
| XX DR WPI; 2002-599511/64.  |     |   |
| XX PT Novel polypeptide for screening inhibitors of non-structural proteases useful as therapeutic agents against hepatitis C virus, comprises full length non-structural protease, or its truncation.  |     |   |
| XX RS Claim 41, Page 59-60; 67pp; English.  |     |   |
| CC The invention relates to an isolated polypeptide consisting of a full-length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal residue amino acid 810 to 905, or having a minimal amino acid sequence from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length NS2/3 protease. Also included are (1) a composition (C) comprising an isolated HCV NS2/3 protease selected from full length NS2/3 protease, or its truncation, or a mutated sequence, where the protease is in a solution comprising a sufficient concentration of lauryldimethylamine oxide (LDLO) to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide appearing as ABG3198; (3) producing (M1) a refolded, inactive HCV NS2/3 protease, involving isolating the protease in the presence of a chaotropic agent, refolding the isolated protease by contacting it with a reducing agent, and LDLO in the presence of reduced concentration of the chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3 protease, involving diluting refolded inactive NS2/3 protease in a medium containing an activation detergent to induce auto-cleavage of the NS2/3 protease; (5) measuring (M3) the auto-cleavage activity of NS2/3 protease, involving incubating the active NS2/3 protease produced by M2 for sufficient time to induce auto-cleavage of NS2/3 protease and produce cleavage products or their fragments, and measuring the presence or absence of uncleaved NS2/3 protease, cleavage products or their fragments and (6) screening a potential inhibitor of auto-cleavage activity of an |     |   |

XX PI Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;  
 XX DR WPI; 2002-599511-64.  
 XX N-PSDB; ABK80406.

PT Novel polypeptide for screening inhibitors of non-structural proteases useful as therapeutic agents against hepatitis C virus, comprises full length non-structural protease, or its truncation.

XX PS Claim 42; Fig 1B; 67pp; English.

CC The invention relates to an isolated polypeptide consisting of a full-length HCV (hepatitis C virus) non-structural (NS2/3) Protease (referred to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal residue amino acid 810 to 906, or having a minimal amino acid sequence from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length NS2/3 protease. Also included are (1) a composition (C) comprising an isolated HCV NS2/3 protease selected from full length NS2/3 protease, or its truncation or a mutated sequence, where the protease is in a solution comprising a sufficient concentration of lauryldiethylamine oxide (LDAO) to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide appearing as ABG32187; (3) producing (M1) a refolded, inactive HCV NS2/3 protease, involving isolating the protease in the presence of a chaotropic agent, refolding the isolated protease by contacting it with a reducing agent, and LDAO in the presence of reduced concentration of the chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3 protease, involving diluting refolded inactive NS2/3 protease in a medium containing an activation detergent to induce auto-cleavage of the NS2/3 protease; (5) measuring (M3) the auto-cleavage activity of NS2/3 protease, involving incubating the active NS2/3 protease produced by M2 for sufficient time to induce auto-cleavage of the NS2/3 protease and produce cleavage products or their fragments, and measuring the presence or absence of uncleaved NS2/3 protease, cleavage products or their fragments ; and (6) screening a potential inhibitor of auto-cleavage activity of an active NS2/3 protease, involving carrying out M3 in the presence of, or absence of the potential inhibitor, comparing the amount of uncleaved NS2/3 protease, cleavage products or their fragments. The protease is useful for detailed biochemical characterisation of the enzymes and in the development of in vitro assays for screening novel inhibitors of NS2/3 Protease which is useful as therapeutic agents against HCV infection (which causes chronic liver disease, cirrhosis and end-stage liver disease). M1 is useful for high level production of protease. The present sequence represents the NS2/3 (810-1206) protein, which has a C-terminal streptavidin tag

CC Sequence 409 AA;

|       |       |  |        |
|-------|-------|--|--------|
| Query | Match | Score  | Length |
| XX    | XX    | 100.0%   | 409    |
| XX    | XX    | 100.0%   | 409;   |
| XX    | XX    | Pred. No. 4.6e-170;  |        |
| XX    | XX    | Matches 352; Conservative 0; Mismatches 0; Indels 0; Gaps 0; |        |
| QY    | 1     | AHLQWIPPLAVNRGGDRDATTILTCVPHPLIDTIKLIAIFGPLMVQAGITKVPFVR     | 60     |
| Db    | 46    | AHQYWPPLNVRGDRDATTILTCVPHPLIDTIKLIAIFGPLMVQAGITKVPFVR        | 105    |
| QY    | 61    | AQGLTRACMVVKRKAAGHYYQMAFKLALITGVMDIPLQWAHAGRLAVAEPV          | 120    |
| Db    | 106   | AQGLTRACMVVKRKAAGHYYQMAFKLALITGVMDIPLQWAHAGRLAVAEPV          | 165    |
| QY    | 121   | IFSDMEVKITTWGADPACCDIISGLPVARRGRERILLGPADNEQGQRKLRAPITASQ    | 180    |
| Db    | 166   | IFSDMEVKITTWGADPACCDIISGLPVARRGRERILLGPADNEQGQRKLRAPITASQ    | 225    |
| QY    | 181   | OTRGGLGCITSLTSDRDKQNGVEGYQVVSATOSFLATCNGVWTWPHGAGKTLAQPK     | 240    |
| Db    | 226   | OTRGGLGCITSLTSDRDKQNGVEGYQVVSATOSFLATCNGVWTWPHGAGKTLAQPK     | 285    |
| QY    | 241   | GPITQMYTNQDVLVQWQPGARSMTPCTCGSSDLYLVRAHVYIPVRGRGSRSILS       | 300    |
| Db    | 286   | GPITQMYTNQDVLVQWQPGARSMTPCTCGSSDLYLVRAHVYIPVRGRGSRSILS       | 345    |
| QY    | 301   | PRPVSIKGSSGGPLCPSPGHAVGIFRAAVCTRGVAKAVDFIPVESMETMR           | 352    |

CC The invention relates to an isolated polypeptide consisting of a full-length HCV (hepatitis C virus) non-structural (NS2/3) Protease (referred to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal residue amino acid 810 to 906, or having a minimal amino acid sequence from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length NS2/3 protease. Also included are (1) a composition (C) comprising an isolated HCV NS2/3 protease selected from full length NS2/3 protease, or its truncation or a mutated sequence, where the protease is in a solution comprising a sufficient concentration of lauryldiethylamine oxide (LDAO) to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide appearing as ABG32187; (3) producing (M1) a refolded, inactive HCV NS2/3 protease, involving isolating the protease in the presence of a chaotropic agent, refolding the isolated protease by contacting it with a reducing agent, and LDAO in the presence of reduced concentration of the chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3 protease, involving diluting refolded inactive NS2/3 protease in a medium containing an activation detergent to induce auto-cleavage of the NS2/3 protease; (5) measuring (M3) the auto-cleavage activity of NS2/3 protease, involving incubating the active NS2/3 protease produced by M2 for sufficient time to induce auto-cleavage of NS2/3 protease and produce cleavage products or their fragments, and measuring the presence or absence of uncleaved NS2/3 protease, cleavage products or their fragments ; and (6) screening a potential inhibitor of auto-cleavage activity of an active NS2/3 protease, involving carrying out M3 in the presence of, or absence of the potential inhibitor, comparing the amount of uncleaved NS2/3 protease, cleavage products or their fragments. The protease is useful for detailed biochemical characterisation of the enzymes and in the development of in vitro assays for screening novel inhibitors of NS2/3 Protease which is useful as therapeutic agents against HCV infection (which causes chronic liver disease, cirrhosis and end-stage liver disease). M1 is useful for high level production of protease. The

DB 346 PRVSYLKKSSGPGCPLSGHVGTFIAVCGRGAVADFTPVEMETMR 397

RESULT 5  
 ABG32187  
 ID ABG32187 standard; protein; 341 AA.  
 XX AC ABG32187;  
 XX DT 05-NOV-2002 (first entry)  
 XX HCV protease NS2/3 truncation mutant 866-1206.  
 DB KW HCV; enzyme; protease; NS2/3 (866-1206); hepatitis C virus infection; chronic liver disease; cirrhosis; end-stage liver disease; viricide; hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO; chaotropic agent; mutant; mutein.  
 XX OS Hepatitis C virus.  
 XX Synthetic.  
 XX WO200248375-A2.  
 XX 15-DEC-2000; 2000US-0256031P.  
 XX PA (BOEHRINGER INGELHEIM CANADA LTD.  
 XX Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A,  
 XX DR WPI; 2002-599511-64.  
 XX PS Novel polypeptide for screening inhibitors of non-structural proteases useful as therapeutic agents against hepatitis C virus, comprises full length non-structural protease, or its truncation.

CC The invention relates to an isolated polypeptide consisting of a full-length HCV (hepatitis C virus) non-structural (NS2/3) protease (referred to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal residue amino acid 810 to 906, or having a minimal amino acid sequence from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length NS2/3 protease. Also included are (1) a composition (C) comprising an isolated HCV NS2/3 protease selected from full length NS2/3 protease, or its truncation or a mutated sequence, where the protease is in a solution comprising a sufficient concentration of lauryldiethylamine oxide (LDAO) to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide appearing as ABG32187; (3) producing (M1) a refolded, inactive HCV NS2/3 protease, involving isolating the protease in the presence of a chaotropic agent, refolding the isolated protease by contacting it with a reducing agent, and LDAO in the presence of reduced concentration of the chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3 protease, involving diluting refolded inactive NS2/3 protease in a medium containing an activation detergent to induce auto-cleavage of the NS2/3 protease; (5) measuring (M3) the auto-cleavage activity of NS2/3 protease, involving incubating the active NS2/3 protease produced by M2 for sufficient time to induce auto-cleavage of NS2/3 protease and produce cleavage products or their fragments, and measuring the presence or absence of uncleaved NS2/3 protease, cleavage products or their fragments ; and (6) screening a potential inhibitor of auto-cleavage activity of an active NS2/3 protease, involving carrying out M3 in the presence of, or absence of the potential inhibitor, comparing the amount of uncleaved NS2/3 protease, cleavage products or their fragments. The protease is useful for detailed biochemical characterisation of the enzymes and in the development of in vitro assays for screening novel inhibitors of NS2/3 Protease which is useful as therapeutic agents against HCV infection (which causes chronic liver disease, cirrhosis and end-stage liver disease). M1 is useful for high level production of protease. The



|  |  |  |  |          |
|--|--|--|--|----------|
| XX   |  | PT   | Region   | 325..327 |
| DR   |  | PT   | /label= N-linked glycosylation site  |          |
| N-PSDB; AAQ80498.  |  | PT   | 417..419   |          |
| XX   |  | Region   | /label= N-linked glycosylation site  |          |
| New HCV-originated proteinase active substance - used for site-specific cleavage by an intermolecular reaction and the purification thereof.   |  | PT   | 423..425   |          |
| PT   |  | Region   | /label= N-linked glycosylation site  |          |
| XX   |  | PT   | 430..432   |          |
| PS Disclosure; Page 10-19; 23pp; Japanese.   |  | Region   | /label= N-linked glycosylation site  |          |
| CC   |  | PT   | 448..450   |          |
| This protein from HCV (hepatitis C virus) (encoded by AAQ80498) is cleaved between amino acids 2419 and 2420, by a new serine protease, contg. the sequence of AAR8621. The proteinase is purified as a fused product with the dihydrofolate reductase protein by using a methotrexate column. It can be used for the development of an inhibitor for HCV proteinase. (Updated on 16-OCT-2003 to standardise OS field) |  | Region   | /label= N-linked glycosylation site  |          |
| XX   |  | PT   | 532..534   |          |
| SQ Sequence 3010 AA;   |  | Region   | /label= N-linked glycosylation site  |          |
| Query Match 95.0%; Score 1768; DB 2; Length 3010;  |  | PT   | 556..558   |          |
| Best Local Similarity 94.3%; Pred. No. 1.5e-161;   |  | Region   | /label= N-linked glycosylation site  |          |
| Matches 332; Conservative 12; Mismatches 8; Indels 0; Gaps 0;  |  | PT   | 576..578   |          |
| QY 1 AHLQWVPPNPPVNGGRDATTILTCAVHPELIFDTKLIAIFGPMVQAGITKVPFVR 60  |  | Region   | /label= N-linked glycosylation site  |          |
| Db 855 AHLQWVPPNPPVNGGRDATTILTCAVHPELIFDTKLIAIFGPMVQAGITKVPFVR 914   |  | PT   | 623..625   |          |
| QY 61 AGQLRACMLVRKAGGHVYOMAFMKLAALTGTYYDHLTPDWAHGLDVLAVFV 120  |  | Region   | /label= N-linked glycosylation site  |          |
| Db 915 AGQLRACMLVRKAGGHVYOMAFMKLAALTGTYYDHLTPDWAHGLDVLAVFV 974   |  | PT   | 645..647   |          |
| QY 121 IFSDNEVKITIWGADTAACCDIISGLPVSAARRGRBILGPADNFEQGMRLAPITASQ 180   |  | Region   | /label= N-linked glycosylation site  |          |
| Db 121 IFSDNEVKITIWGADTAACCDIISGLPVSAARRGRBILGPADNFEQGMRLAPITASQ 2041..2043  |  | PT   | 663..665   |          |
| QY 975 VFSMETKLITWGADTAACCDIISGLPVSAARRGRBILGPADNFEQGMRLAPITASQ 1034   |  | Region   | /label= N-linked glycosylation site  |          |
| Db 181 QTRGLIGCITSLTDQDLYGWPAFPGARSMTCTCGSSDLVLTTRHADVTPVRRDPSRSILLS 240   |  | PT   | 682..684   |          |
| QY 1035 QTRGLIGCITSLTDQDLYGWPAFPGARSMTCTCGSSDLVLTTRHADVTPVRRDPSRSILLS 974  |  | Region   | /label= N-linked glycosylation site  |          |
| Db 241 GPITOMYTNDQDLYGWPAFPGARSMTCTCGSSDLVLTTRHADVTPVRRDPSRSILLS 300   |  | PT   | 707..709   |          |
| QY 1095 GPITOMYTNDQDLYGWPAFPGARSMTCTCGSSDLVLTTRHADVTPVRRDPSRSILLS 1154   |  | Region   | /label= N-linked glycosylation site  |          |
| Db 301 PRPSYLKQSSGGPCLPCEPGHGVGIFRAVCTRVAKAVDFIVVESMTTMR 352   |  | PT   | 727..729   |          |
| QY 1155 PRPSYLKQSSGGPCLPCEPGHGVGIFRAVCTRVAKAVDFIVVESMTTMR 1206   |  | Region   | /label= N-linked glycosylation site  |          |
| Db   |  | PT   | 747..749   |          |
| RESULT 8   |  | DN   | JP05319583-A.  |          |
| ID AAR8864   |  | XX   |  |          |
| ID AAR8864 standard; protein; 3010 AA.   |  | PD   | 22-NOW-1994.   |          |
| XX   |  | XX   |  |          |
| AC AAR8864;  |  | PR   | 18-SEP-1992; 92JP-00249241.  |          |
| XX   |  | XX   |  |          |
| DT 06-DEC-1995 (first entry)   |  | PR   | 18-SBP-1992; 92JP-00249241.  |          |
| XX   |  | XX   |  |          |
| DE Hepatitis C virus RNA helicase.   |  | CA   | (SOYA-) SOYAKU GIUTSU KENKYUSHO KK.  |          |
| XX   |  | XX   |  |          |
| HE Hepatitis C virus helicase gene in baculovirus - useful for large PT scale prodn. of RNA helicase.  |  | CC   | WPI: 1995-040330/06.   |          |
| KW baculovirus; recombinant production.  |  | CC   | DR N-PSDB; AAQ81559.   |          |
| XX   |  | XX   |  |          |
| OS Hepatitis C virus.  |  | CC   | AAQ81559 encodes AAR8864 hepatitis C virus (HCV) RNA helicase. The DNA was used in the construction of an expression vector, which was used to transform a baculovirus host. The transformed baculovirus could then be used for the recombinant prodn. of HCV RNA helicase |          |
| XX   |  | CC   |  |          |
| Key Location/Qualifiers  |  | CC   | Sequence 3010 AA;  |          |
| FT Region 166..198 /label= N-linked glycosylation site   |  | Query Match 95.9%; Score 1767; DB 2; Length 3010;                    |  |          |
| FT Region 209..211 /label= N-linked glycosylation site   |  | Best Local Similarity 94.3%; Pred. No. 1.5e-161;                     |  |          |
| FT Region 234..236 /label= N-linked glycosylation site   |  | Matches 332; Conservative 12; Mismatches 8; Indels 0; Gaps 0;        |  |          |
| FT Region 250..252 /label= N-linked glycosylation site   |  | QY 1 AHLQWVPPNPPVNGGRDATTILTCAVHPELIFDTKLIAIFGPMVQAGITKVPFVR 60      |  |          |
| FT Region 305..307 /label= N-linked glycosylation site   |  | Db 855 AHLQWVPPNPPVNGGRDATTILTCAVHPELIFDTKLIAIFGPMVQAGITKVPFVR 914   |  |          |
| FT Region 305..307 /label= N-linked glycosylation site   |  | QY 61 AGQLRACMLVRKAGGHVYOMAFMKLAALTGTYYDHLTPDWAHGLDVLAVFV 120        |  |          |
| FT Region 305..307 /label= N-linked glycosylation site   |  | Db 915 AGQLRACMLVRKAGGHVYOMAFMKLAALTGTYYDHLTPDWAHGLDVLAVFV 974       |  |          |
| FT Region 305..307 /label= N-linked glycosylation site   |  | QY 121 IFSDNEVKITIWGADTAACCDIISGLPVSAARRGRBILGPADNFEQGMRLAPITASQ 180 |  |          |
| FT Region 305..307 /label= N-linked glycosylation site   |  | Db 975 VFSMETKLITWGADTAACCDIISGLPVSAARRGRBILGPADNFEQGMRLAPITASQ 1034 |  |          |

CC the claims of the invention  
 XX  
 SQ Sequence 2201 AA;  
 Query Match 95.9%; Score 1766; DB 5; Length 2201;  
 Best Local Similarity 94.3%; Pred. No. 1.2e-161;  
 Matches 332; Conservative 11; Mismatches 9; Indels 0; Gaps 0;  
 Db 1095 APIQMTNTVDQDLVGWQAPPGARSMTPTCTGSSDLYLVRHADVTPVRRGDSRGSLIS 1154  
 QY 301 PRPVSTKGGGGILCPSGHAVGIFRAVCTGVAKADEPVESETTMR 352  
 Db 1155 PRPISYLKGSSGGPLICPSGHAVGIFRAVCTGVAKAVDFIPVESMETMR 1206  
 RESULT 9  
 ABG30601  
 ID ABG30601 standard; protein; 2201 AA.  
 XX  
 AC ABG30601;  
 XX  
 DT 21-OCT-2002 (first entry)  
 DE Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B mutant #10.  
 XX  
 KW Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;  
 KW cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; mutein.  
 XX  
 OS Hepatitis C virus.  
 XX  
 PH Location/Qualifiers  
 FT Misc-difference 882  
 FT /label= Arg, Lys  
 FT  
 FT Misc-difference 2183  
 /note= "Wild type Met substituted by Thr"  
 PN WO200252015-A2.  
 XX  
 PD 04-JUL-2002.  
 XX  
 PF 20-DEC-2001; 2001WO-CA001843.  
 XX  
 PR 22-DEC-2000; 2000US-0257857P.  
 XX  
 PA (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  
 PI Kukolj, G., Pause, A.;  
 XX  
 DR WPI; 2002-575382/61.  
 XX  
 PT New self-replicating RNA molecules from Hepatitis C virus (HCV), which  
 PT possess enhanced transduction or replication efficiency, useful for  
 evaluating potential inhibitors of HCV replication.  
 XX  
 PS Claim 3; Page; 140PP; English.  
 XX  
 CC The invention describes a self-replicating hepatitis C virus (HCV), which  
 CC polynucleotide molecule comprising a 5'-non translated region (NTR),  
 CC where guanine at position 1 is substituted for adenine, a HCV polyprotein  
 CC region coding for a HCV polyprotein, and a 3'-NTR region. The self-  
 CC replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating  
 CC potential inhibitors of HCV replication. The HCV RNA molecule is also  
 CC useful for efficiently establishing cell culture replication. The self-  
 CC replicating polynucleotide molecule contains a 5'-NTR, where G at  
 CC position 1 is substituted for A, and therefore provides an alternative to  
 CC existing systems comprising a self-replicating HCV RNA molecule that,  
 CC in conjunction with mutations in the HCV non-structural region, such as the  
 CC G1042C/R mutations, transduces and/or replicates with greater  
 CC efficiency. This amino acid sequence represents a mutant of the hepatitis  
 CC virus replicon APGK12 and contains the viral protease NS2/3, protease  
 CC complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note:  
 CC This sequence does not appear in the specification but has been created  
 from the wild type sequence shown in ABG30580 using information given in  
 XX  
 CC the claims of the invention  
 XX  
 SQ Sequence 2201 AA;  
 Query Match 95.9%; Score 1766; DB 5; Length 2201;  
 Best Local Similarity 94.3%; Pred. No. 1.2e-161;  
 Matches 332; Conservative 11; Mismatches 9; Indels 0; Gaps 0;  
 Db 46 ALIQWVPIPVLNRGRDAVILTCIHPPELJFPTKILAILGPMLQAGITKPYFR 105  
 QY 61 AQGLIRACMLYTKRKAAGGYYQMAFKLALGTGTYWDPLQDMAHAGIRDLAYAPEV 120  
 Db 106 ARGILRACMLYTKRKAAGGYYQMAFKLALGTGTYWDPLQDMAHAGIRDLAYAPEV 165  
 QY 121 IFSDMUVLWILNGACTAACGDIISLIPVSARGRSRELLSGADNEQQGMLAPTAYSQ 180  
 Db 166 VFSDMETKIVTGADTAACGDIILGPILPVSAARRGRETHLPGQGRHLAPTAYSQ 225  
 QY 181 QTRGLGCITSLSITGRDKNQVEGVQVSTATQFLATCYNGVCMVFHAGSKTLAGPK 240  
 Db 226 QTRGLGCITSLSITGRDKNQVEGVQVSTATQFLATCYNGVCMVFHAGSKTLAGPK 285  
 QY 241 APIQMTNTVDQDLVGWQAPPGARSMTPTCTGSSDLYLVRHADVTPVRRGDSRGSLIS 300  
 Db 286 APIQMTNTVDQDLVGWQAPPGARSMTPTCTGSSDLYLVRHADVTPVRRGDSRGSLIS 345  
 QY 301 PRPVSTKGGGGILCPSGHAVGIFRAVCTGVAKADEPVESETTMR 352  
 Db 346 PRPISYLKGSSGGPLICPSGHAVGIFRAVCTGVAKAVDFIPVESMETMR 397  
 RESULT 10  
 ABG30591  
 ID ABG30591 standard; protein; 2201 AA.  
 XX  
 AC ABG30591;  
 XX  
 DT 21-OCT-2002 (first entry)  
 DE Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B mutant #3.  
 XX  
 PR Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;  
 KW cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; mutein.  
 XX  
 OS Hepatitis C virus.  
 XX  
 PH Location/Qualifiers  
 FT Misc-difference 751  
 FT /note= "Wild type Ser substituted by Gly"  
 FT Misc-difference 882  
 FT /label= Arg, Lys  
 PN WO200252015-A2.  
 XX  
 PD 04-JUL-2002.  
 XX  
 PF 20-DEC-2001; 2001WO-CA001843.  
 XX  
 PR 22-DEC-2000; 2000US-0257857P.  
 XX  
 PA (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  
 XX  
 PI Kukolj, G., Pause, A.;  
 XX  
 DR WPI; 2002-575382/61.  
 XX  
 PT New self-replicating RNA molecules from Hepatitis C virus (HCV), which  
 PT possess enhanced transduction or replication efficiency, useful for  
 evaluating potential inhibitors of HCV replication.

PS Claim 3; Page; 140pp; English.

XX WO200252015-A2.

CC The invention describes a self-replicating hepatitis C virus (HCV) polynucleotide molecule comprising a 5'-non translated region (NTR), where guanine at position 1 is substituted for adenine, a HCV polyprotein region coding for a HCV polyprotein; and a 3'-NTR region. The self-replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating potential inhibitors of HCV replication. The HCV RNA molecule is also useful for efficiently establishing cell culture replication. The self-replicating polynucleotide molecule contains a 5'-NTR, where G at position 1 is substituted for A, and therefore provides an alternative to existing systems comprising a self-replicating HCV RNA molecule that, in conjunction with mutations in the HCV non-structural region, such as the G12042/C/R mutations, transduces and/or replicates with greater efficiency. This amino acid sequence represents a mutant of the hepatitis C virus replicon ARGK12 and contains the viral protease NS2/3, protease complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note: This sequence does not appear in the specification but has been created from the wild type sequence shown in ABG30580 using information given in the claims of the invention

CC XX

CC SQ

CC Sequence 2201 AA;

CC Query Match 95.9%; Score 1766; DB 5; Length 2201;

CC Best Local Similarity 94.3%; Pred. No. 1-161; Mismatches 9; Indels 0; Gaps 0;

CC Matches 332; Conservative 11; MisMatches 9; Indels 0; Gaps 0;

CC Qy 1 AHLQWIPPLNVRRGGRDAVILTCATIHPBLIFITKILLAILGLPMVHQAGITKVYFVR 60

CC Db 46 AHLQWIPPLNVRRGGRDAVILTCATIHPBLIFITKILLAILGLPMVHQAGITKVYFVR 105

CC Qy 61 AQLRTRACMVVKRAAGGHTYQAMFKMLAALTGTYYWYHPTIOPDWAHGLDLAVAPVY 120

CC Db 105 AQLRTRACMVVKYAGGHVYQAMFKMLAALTGTYYWYHPTIOPDWAHGLDLAVAPVY 165

CC Qy 121 IFSPMEVKITIWGAIDTAACGDIISGLPVSAARRGRELILGPADNEFGQWRLLAPIYASQ 180

CC Db 166 VFSDEMETKIVTWGAIDTAACGDIISGLPVSAARRGRELILGPADNEFGQWRLLAPIYASQ 225

CC Qy 181 QTRGLIGCITSLTRGLDKNOVEGEVQVNSTATOSFLATCVNGCWTWFGAGSKTLAGPK 240

CC Db 226 QTRGLIGCITSLTRGLDKNOVEGEVQVNSTATOSFLATCVNGCWTWFGAGSKTLAGPK 285

CC Qy 241 GPITQMYTNTDQDLYVGWQAPPGARSMTCTCSDDLYVTRHADVTRRGSRSLLS 300

CC Db 286 GPITQMYTNTDQDLYVGWQAPPGARSILPCTCGSSDLVYLTTRHADVTRRGSRSLLS 345

CC Qy 301 PRPVSYLKGSGGPUCPSGHAVGIFRAAVCTRGVAKAVDFPVVESMFTMR 352

CC Db 346 PRPVSYLKGSGGPUCPSGHAVGIFRAAVCTRGVAKAVDFPVVESMFTMR 397

CC RESULT 11

CC ABG30600

CC ID ABG30600 standard; protein; 2201 AA.

CC XX

CC AC ABG30600;

CC XX 21-OCT-2002 (first entry)

CC DE Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B mutant #9.

CC XX Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor; cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; mutein.

CC OS Hepatitis C virus.

CC OS Synthetic.

CC XX

CC Key F8

CC FT Misc-difference 882

CC FT Misc-difference 1357

CC FT /note= "Wild type Pro substituted by Leu"

XX PN WO200252015-A2.

XX PD 04-JUL-2002.

XX DT 20-DEC-2001; 2001WO-CA001843.

XX FR 22-DEC-2000; 2000US-0257857B.

XX PA (BOEHRINGER INGELHEIM CANADA LTD.

XX PI Kukolj, G., Pause A;

XX DR WPI; 2002-575382/61.

XX PT New self-replicating RNA molecules from Hepatitis C virus (HCV), which possess enhanced transduction or replication efficiency, useful for evaluating potential inhibitors of HCV replication.

XX PT

CC PG Claim 3; Page; 140pp; English.

CC XX The invention describes a self-replicating hepatitis C virus (HCV) polynucleotide molecule comprising a 5'-non translated region (NTR), where guanine at position 1 is substituted for adenine, a HCV polyprotein region coding for a HCV polyprotein; and a 3'-NTR region. The self-replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating potential inhibitors of HCV replication. The HCV RNA molecule is also useful for efficiently establishing cell culture replication. The self-replicating polynucleotide molecule contains a 5'-NTR, where G at position 1 is substituted for A, and therefore provides an alternative to existing systems comprising a self-replicating HCV RNA molecule that, in conjunction with mutations in the HCV non-structural region, such as the G12042/C/R mutations, transduces and/or replicates with greater efficiency. This amino acid sequence represents a mutant of the hepatitis C virus replicon ARGK12 and contains the viral protease NS2/3, protease complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note: This sequence does not appear in the specification but has been created from the wild type sequence shown in ABG30580 using information given in the claims of the invention

CC XX

CC SQ

CC Sequence 2201 AA;

CC Query Match 95.9%; Score 1766; DB 5; Length 2201;

CC Best Local Similarity 94.3%; Pred. No. 1-161; Mismatches 9; Indels 0; Gaps 0;

CC Matches 332; Conservative 11; MisMatches 9; Indels 0; Gaps 0;

CC Qy 1 AHLQWIPPLNVRRGGRDAVILTCATIHPBLIFITKILLAILGLPMVHQAGITKVYFVR 60

CC Db 46 AHLQWIPPLNVRRGGRDAVILTCATIHPBLIFITKILLAILGLPMVHQAGITKVYFVR 105

CC Qy 61 AQLRTRACMVVKRAAGGHTYQAMFKMLAALTGTYYWYHPTIOPDWAHGLDLAVAPVY 120

CC Db 105 AQLRTRACMVVKYAGGHVYQAMFKMLAALTGTYYWYHPTIOPDWAHGLDLAVAPVY 165

CC Qy 121 IFSPMEVKITIWGAIDTAACGDIISGLPVSAARRGRELILGPADNEFGQWRLLAPIYASQ 180

CC Db 166 VFSDEMETKIVTWGAIDTAACGDIISGLPVSAARRGRELILGPADNEFGQWRLLAPIYASQ 225

CC Qy 181 QTRGLIGCITSLTRGLDKNOVEGEVQVNSTATOSFLATCVNGCWTWFGAGSKTLAGPK 240

CC Db 226 QTRGLIGCITSLTRGLDKNOVEGEVQVNSTATOSFLATCVNGCWTWFGAGSKTLAGPK 285

CC Qy 241 GPITQMYTNTDQDLYVGWQAPPGARSMTCTCSDDLYVTRHADVTRRGSRSLLS 300

CC Db 286 GPITQMYTNTDQDLYVGWQAPPGARSILPCTCGSSDLVYLTTRHADVTRRGSRSLLS 345

CC Qy 301 PRPVSYLKGSGGPUCPSGHAVGIFRAAVCTRGVAKAVDFPVVESMFTMR 352

CC Db 346 PRPVSYLKGSGGPUCPSGHAVGIFRAAVCTRGVAKAVDFPVVESMFTMR 397

CC RESULT 12

CC ABG30581

ID ABG30581 standard; protein; 2201 AA.  
 XX  
 AC ABG30581;  
 XX  
 DT 21-OCT-2002 (first entry)  
 XX Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B #1.  
 KW self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;  
 XX cell culture replication; NS2/3; NS3/4; NS3; NS5B.  
 OS Hepatitis C virus.  
 XX  
 PN WO200252015-A2.  
 XX  
 PD 04-JUL-2002.  
 PF 20-DEC-2001; 2001WO-CA001843.  
 XX PR 22-DEC-2000; 2000US-0257857P.  
 XX PA (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  
 XX PT Kukolj G, Pause A;  
 XX DR WPI; 2002-575382/61.  
 DR N-PSDB; ABK08573.  
 XX PT New self-replicating RNA molecules from Hepatitis C virus (HCV), which  
 possess enhanced transduction or replication efficiency, useful for  
 evaluating potential inhibitors of HCV replication.  
 XX PS Disclosure; Page 49-58; 140pp; English.  
 XX CC The invention describes a self-replicating hepatitis C virus (HCV)  
 poly nucleotide molecule comprising a 5'-non translated region (NTR),  
 where guanine at position 1 is substituted for adenine, a HCV polyprotein  
 region coding for a HCV polyprotein; and a 3'-NTR region. The self-  
 replicating Hepatitis C Virus (HCV) RNA molecule is useful for evaluating  
 potential inhibitors of HCV replication. The HCV RNA molecule is also  
 useful for efficiently establishing cell culture replication. The self-  
 replicating poly nucleotide molecule contains a 5'-NTR, where G at  
 position 1 is substituted for A, and therefore provides an alternative to  
 existing systems comprising a self-replicating HCV RNA molecule that, in  
 conjunction with mutations in the HCV non-structural region, such as the  
 G(2042)C/R mutations, transduces and/or replicates with greater  
 efficiency. This amino acid sequence is encoded by the hepatitis C virus  
 replicon APGK12 and contains the viral protease NS2/3, protease complex  
 CC NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B  
 CC Sequence 2201 AA;  
 XX SQ Query Match 95.9%; Score 1766; DB 5; Length 2201;  
 Best Local Similarity 94.3%; Pred. No. 1.2e-161; Mismatches 9; Indels 0; Gaps 0;  
 Matches 332; Conservative 11; Nucleotides 9;  
 Qy 1 AHIQWIPPLNVRGRDAILTCAVHPPLIDPTIKLAIARPLMLIQAGITKTVYFVR 60  
 Db 46 AHLQWIPPLNVRGRDAILTCAIPELIFTIKLAIACGPMVLQAGITKTVYFVR 105  
 Qy 61 AQGLIRACMVLRKAAAGHIVQMAFMKLAALTGTIVYDHLPLODWAHAGJRLAVAVEPV 120  
 Db 106 AHLGLIRACMVLRKAAAGHIVQMAFMKLAALTGTIVYDHLPLODWAHAGJRLAVAVEPV 165  
 Qy 121 IFSDMEVKITIWGADTAAGDIISGLPV/SARGRBILIGPAMNPEGCOWRLLAPITASQ 180  
 Db 165 VFSMETKVTWGAATGADTAAGDIISGLPV/SARGRBILIGPAMNPEGCOWRLLAPITASQ 225  
 Qy 181 QTRGLIGCITSLSTGRDKNOVEGEVQMVSTATOSFLATCVNGYCWTWFGAGSKTLAGPK 240  
 Db 226 QTRGLIGCITSLSTGRDKNOVEGEVQMVSTATOSFLATCVNGYCWTWFGAGSKTLAGPK 285  
 XX 241 GPITQMYTNVDQDLVGWQAPPGARSMTPTCGSSDLYLVTRADVIPVRRDSRGSSLS 300

RESULT 13  
 ID ABG30593 standard; protein; 2201 AA.  
 XX  
 AC ABG30593;  
 XX  
 DT 21-OCT-2002 (first entry)  
 XX Hepatitis C virus NS2/3, NS3 and NS5B mutant #4.  
 DE Hepatitis C virus NS2/3, NS3 and NS5B mutant #4.  
 XX Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;  
 KW cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; mutein.  
 XX PA (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  
 OS Hepatitis C virus.  
 OS Synthetic.  
 XX PT Key Location/Qualifiers  
 FT Misc-difference 882  
 FT /label= Arg, Lys  
 FT Misc-difference 892  
 FT /note= "Wild type Leu substituted by Phe"  
 XX PT PN 20-DEC-2001; 2001WO-CA001843.  
 XX PR PN WO200252015-A2.  
 XX PD 04-JUL-2002.  
 XX PT Kukolj G, Pause A;  
 XX PR 22-DEC-2000; 2000US-0257857P.  
 XX PA (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  
 PT DR WPI; 2002-575382/61.  
 XX PT New self-replicating RNA molecules from Hepatitis C virus (HCV), which  
 possess enhanced transduction or replication efficiency, useful for  
 evaluating potential inhibitors of HCV replication.  
 XX PS Claim 3; Page; 140pp; English.  
 XX CC The invention describes a self-replicating hepatitis C virus (HCV)  
 poly nucleotide molecule comprising a 5'-non translated region (NTR),  
 where guanine at position 1 is substituted for adenine, a HCV polyprotein  
 region coding for a HCV polyprotein; and a 3'-NTR region. The self-  
 replicating Hepatitis C Virus (HCV) RNA molecule is useful for evaluating  
 potential inhibitors of HCV replication. The HCV RNA molecule is also  
 useful for efficiently establishing cell culture replication. The self-  
 replicating poly nucleotide molecule contains a 5'-NTR, where G at  
 position 1 is substituted for A, and therefore provides an alternative to  
 existing systems comprising a self-replicating HCV RNA molecule that, in  
 conjunction with mutations in the HCV non-structural region, such as the  
 G(2042)C/R mutations, transduces and/or replicates with greater  
 efficiency. This amino acid sequence represents a mutant of the hepatitis  
 C virus replicon APGK12 and contains the viral protease NS2/3, protease  
 complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note:  
 CC This sequence does not appear in the specification but has been created  
 CC from the wild type sequence shown in ABG30580 using information given in  
 CC the claims of the invention  
 XX Sequence 2201 AA;

Query Match 95.9%; Score 1766; DB 5; Length 2201;  
 Best Local Similarity 94.3%; Pred. No. 1.2e-161;

DB 286 GPITQMYTNVDQDLVGWQAPPGARSMTPTCGSSDLYLVTRADVIPVRRDSRGSSLS 345  
 Qy 301 PRPVSYLKGSQGPLCPSGHAVGIFRAVCTRGVAKAVDVFVPVESMETMR 352  
 DB 346 PRPVSYLKGSQGPLCPSGHAVGIFRAVCTRGVAKAVDVFVPVESMETMR 397

Matches 332; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

QY 1 AHLQWIPPLNVRGRDAILLTCAVHFLIFDITKULLAIFGPMIQLAGITKIVYFVR 60  
Db 46 AHLQWIPPLNVRGRDAILLTCAVHFLIFDITKULLAIFGPMIQLAGITKIVYFVR 105

QY 61 AQLTRACMVRKAAGGYQMAALKLALLTGTIVYVYELTPQDWAHGLRDLAVAEPV 120  
Db 106 AHLTRACMVRKAAGGYQMAALKLALLTGTIVYVYELTPQDWAHGLRDLAVAEPV 165

QY 121 IFSDMEVKITWTGADTAACDDILGLPVSARRGREIILGPAPNFEQGWRLLAPIYSQ 180  
Db 166 VFSDEMETKVITWTGADTAACDDILGLPVSARRGREIILGPAPNFEQGWRLLAPIYSQ 225

QY 181 QTRGLIGCITSLSGRDKQVEGVQVSTATOSFLATCVNGVCWTVFHGSKTLAGPK 240  
Db 226 QTRGLIGCITSLSGRDRNQVEGVQVSTATOSFLATCVNGVCWTVFHGSKTLAGPK 285

QY 241 GPITQMYNTWDQDGWQAPPGRASMTCTCGSDLYNTRHADVIPRRGDSRSILS 300  
Db 285 GPITQMYNTWDQDGWQAPPGRASLTPTCTCGSDLYNTRHADVIPRRGDSRSILS 345

QY 301 PRPVSYLKSSGGPLICPSGHAVGIFRAAVCTRGVAKAVDFPVVESMETMR 352  
Db 346 PRPVSYLKSSGGPLICPSGHAVGIFRAAVCTRGVAKAVDFPVVESMETMR 397

RESULT 14

ID ABG30582 standard; protein: 2201 AA.

XX AC ABG30582;

XX DT 21-OCT-2002 (first entry)

XX DE Hepatitis C virus NS2/3, NS3/4, NS3 and NSSB #2.

XX KW Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor; cell culture replication; NS2/3; NS3/4; NS3; NSSB; mutant; mutein.

XX OS Hepatitis C virus.

XX PH Key Location/Qualifiers

FT Misc-difference 882  
/note= "Wild type Lys substituted by Lys or Arg"

FT Misc-difference 1233  
/note= "Wild type Gly substituted by Cys"

PN WO200252015-A2.

XX PD 04-JUL-2002.

XX PT 20-DEC-2001; 2001WO-CA001843.

XX PR 22-DEC-2000; 2000US-0257857P.

PA (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.

XX PI Kukolj G, Pause A;

XX DR WPI; 2002-57382/61..  
DR N-PSDB; ABK86574.

XX PT New self-replicating RNA molecules from Hepatitis C virus (HCV), which possess enhanced transduction or replication efficiency, useful for evaluating potential inhibitors of HCV replication.

XX PS Disclosure; Page 59-59; 140PP; English.

CC The invention describes a self-replicating RNA molecule comprising a 5'-non translated region (NTR), where guanine at position 1 is substituted for adenine, a HCV polyprotein region coding for a HCV polyprotein; and a 3'-NTR region. The self-

CC sequence has been created from replicon ARPK12 shown in ABG30581

XX SQ Sequence 2201 AA:

Query Match 95.9%; Score 1766; DB 5; Length 2201;  
Best Local Similarity 94.3%; Pred. No. 1.2e-161; Mismatches 9; Indels 0; Gaps 0;

Matches 332; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

QY 1 AHLQWIPPLNVRGRDAILLTCAVHFLIFDITKULLAIFGPMIQLAGITKIVYFVR 60  
Db 46 AHLQWIPPLNVRGRDAILLTCAVHFLIFDITKULLAIFGPMIQLAGITKIVYFVR 105

QY 61 AQLTRACMVRKAAGGYQMAALKLALLTGTIVYVYELTPQDWAHGLRDLAVAEPV 120  
Db 106 AHLTRACMVRKAAGGYQMAALKLALLTGTIVYVYELTPQDWAHGLRDLAVAEPV 165

QY 121 IFSDMEVKITWTGADTAACDDILGLPVSARRGREIILGPAPNFEQGWRLLAPIYSQ 180  
Db 166 VFSDEMETKVITWTGADTAACDDILGLPVSARRGREIILGPAPNFEQGWRLLAPIYSQ 225

QY 181 QTRGLIGCITSLSGRDKQVEGVQVSTATOSFLATCVNGVCWTVFHGSKTLAGPK 240  
Db 226 QTRGLIGCITSLSGRDRNQVEGVQVSTATOSFLATCVNGVCWTVFHGSKTLAGPK 285

QY 241 GPITQMYNTWDQDGWQAPPGRASMTCTCGSDLYNTRHADVIPRRGDSRSILS 300  
Db 285 GPITQMYNTWDQDGWQAPPGRASLTPTCTCGSDLYNTRHADVIPRRGDSRSILS 345

QY 301 PRPVSYLKSSGGPLICPSGHAVGIFRAAVCTRGVAKAVDFPVVESMETMR 352  
Db 346 PRPVSYLKSSGGPLICPSGHAVGIFRAAVCTRGVAKAVDFPVVESMETMR 397

RESULT 15

ID ABG30580 standard; protein: 2201 AA.

XX AC ABG30580;

XX DT 21-OCT-2002 (first entry)

XX DE Hepatitis C virus NS2/3, NS3/4, NS3 and NSSB #9.

XX KW Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor; cell culture replication; NS2/3; NS3/4; NS3; NSSB.

XX OS Hepatitis C virus.

XX PH Key Location/Qualifiers

FT Misc-difference 882  
/note= "Encoded by ARG"

XX PN WO200252015-A2.

XX PD 04-JUL-2002.

XX PT 20-DEC-2001; 2001WO-CA001843.

XX PR 22-DEC-2000; 2000US-0257857P.

PA (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.

PT Kukolj G, Pause A;  
 XX WPI; 2002-575382/61.  
 DR

PT New self-replicating RNA molecules from Hepatitis C virus (HCV), which  
 possess enhanced transduction or replication efficiency, useful for  
 evaluating potential inhibitors of HCV replication.

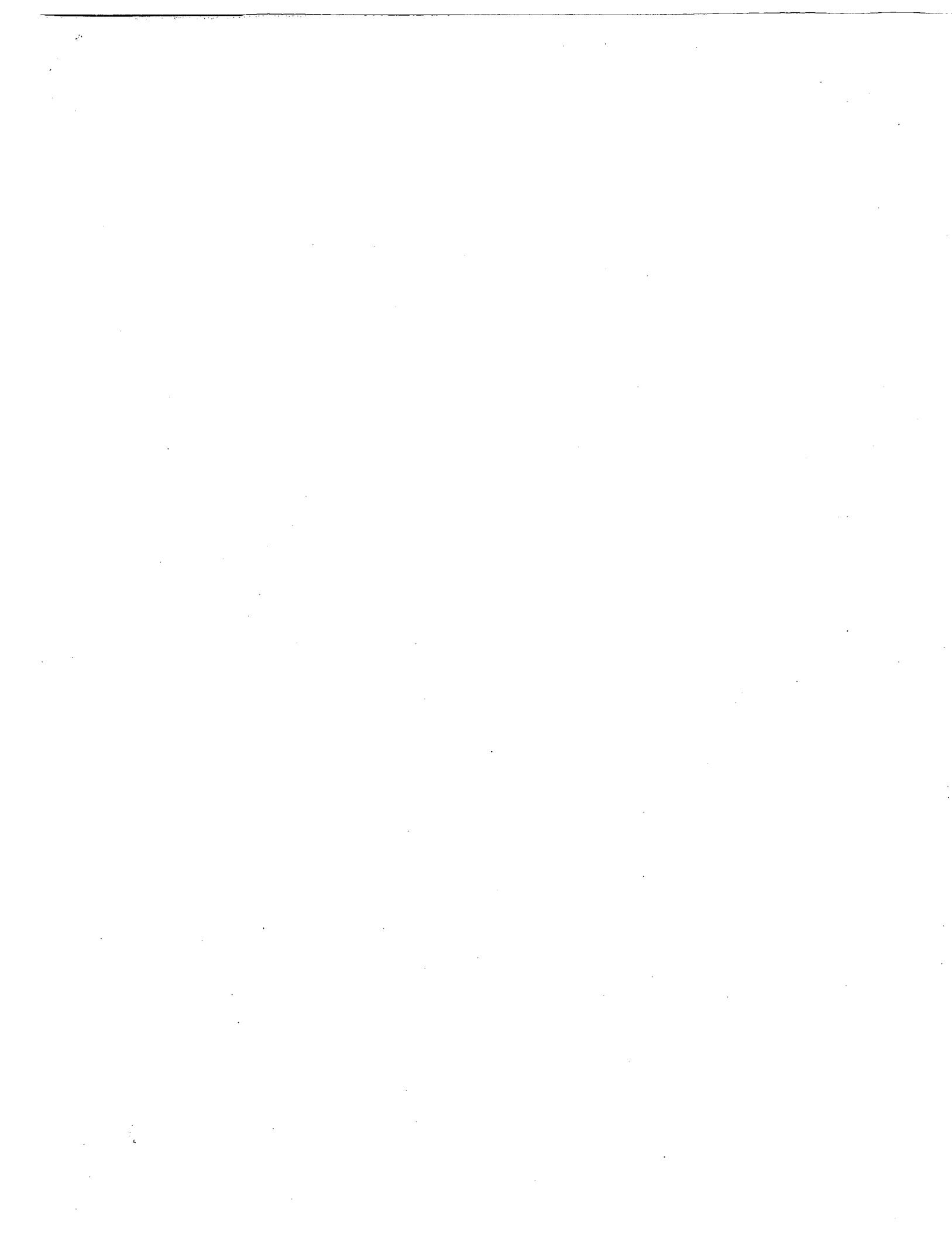
XX Disclosure; Page 69-74; 140pp; English.

The invention describes a self-replicating hepatitis C virus (HCV) polyribonucleotide molecule comprising a 5'-non translated region (NTR), where guanine at position 1 is substituted for adenine, a HCV polyprotein region coding for a HCV polyprotein, and a 3'-NTR-region. The self-replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating potential inhibitors of HCV replication. The HCV RNA molecule is also useful for efficiently establishing cell culture replication. The self-replicating Polyribonucleotide molecule contains a 5'-NTR, where G at position 1 is substituted for A, and therefore provides an alternative to existing systems comprising a self-replicating HCV RNA molecule that, in conjunction with mutations in the HCV non-structural region, such as the G-2042/C/R mutations, transduces and/or replicates with greater efficiency. This amino acid sequence is encoded by the hepatitis C virus replicon APCK12 and contains the viral protease NS2/3, protease complex CC and RNA-dependent RNA polymerase NS5B CC NS3/4, helicase NS3 and XX Sequence 2201 AA;

Query Match 95.9%; Score 1766; DB 5; Length 2201;  
 Best Local Similarity 94.3%; Pred No 1.2e-151; Mismatches 9; Indels 0; Gaps 0;  
 Matches 332; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

Qy 1 AHIQWIPINVGGRDAITLTCAVHIELPITKLLAIGPLMVQAGITTKVPFVVR 60  
 Db 46 AHIQWIPINVGGRDAITLTCAVHIELPITKLLAIGPLMVQAGITTKVPFVVR 105  
 Qy 61 AQQGIRACMLVKRAGGHYVQAFMKLALITGTYVYDILTPLODWAHLGLRLAVAPEV 120  
 Db 106 AHHGLRACMLVKRAGGHYVQAFMKLALITGTYVYDILTPLODWAHLGLRLAVAPEV 165  
 Qy 121 IFSMEMVKITIWGAIDAAQGDIISGJFSPSARRREILGPNPEGGWRLAPIYSQ 180  
 Db 166 VFSMETKVITWGADTAAGGDIILGPISARERREIHGPALSLEGQWRJLAPIYSQ 225  
 Qy 181 QTRQGLGCITTSITGRDKNQVEGVSVQVSTATQSFLATCVNGSVCWTVPHGASKTLGPK 240  
 Db 225 QTRQGLGCITTSITGRDNQVEGVSVQVSTATQSFLATCVNGSVCWTVPHGASKTLGPK 285  
 Qy 241 GRPTOMYNTVDQDILVGWDRAPPGRSMTCCTCSDLYVTRADIVPYRRGDSRSILLS 300  
 Db 286 GPITQMYNTVDQDILVGWDRAPPGRSLTCTCGSSDLYVTRADIVPYRRGDSRSILLS 345  
 Qy 301 PRPVSYLKGSSEPLCLCPSGHAGVIFRAVCTRGVAKDFPVESETTMR 352  
 Db 346 PRPVSYLKGSSEPLCLCPSGHAGVIFRAVCTRGVAKDFPVESETTMR 397

Search completed: May 6, 2004, 09:30:47  
 Job time : 48.912 secs



GenCore version 5.1.6  
 Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on:

May 6, 2004, 09:22:36 ; search time 11:4639 Seconds

(without alignments)  
 2953.573 Million cell updates/sec

Title: US-10-650-585-13

Perfect score: 1842

Sequence: AHLQWVPPPLNVGRGDAILLTCAVHPPELFIDITKLLAIFGPMLVLOGITKVYPFR

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 78:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query | Length   | DB ID               | Description                                       |
|------------|-------|-------|----------|---------------------|---|
| 1          | 1773  | 96.3  | 3010     | 1 GNVWCJ            | genome polyprotein - hepatitis C virus (strain J) |
| 2          | 1766  | 95.9  | 3010     | 1 A45573            | genomic polyprotein                               |
| 3          | 1749  | 95.0  | 3010     | 1 GNTWTC            | genomic polyprotein                               |
| 4          | 1717  | 93.3  | 3010     | 1 S18303            | genomic polyprotein                               |
| 5          | 1699  | 92.2  | 3010     | 1 GNWVC             | genomic polyprotein                               |
| 6          | 1616  | 87.7  | 3011     | 1 GNTWTC            | genomic polyprotein                               |
| 7          | 87.4  | 3011  | 1 S40770 | genomic polyprotein |   |
| 8          | 1605  | 87.1  | 3011     | 1 GNTWCH            | genomic polyprotein                               |
| 9          | 1401  | 76.1  | 3014     | 1 JC5620            | genomic polyprotein                               |
| 10         | 1303  | 70.7  | 3033     | 1 GNTWTC            | genomic polyprotein                               |
| 11         | 1301  | 70.6  | 3033     | 1 QJ01303           | genomic polyprotein                               |
| 12         | 408.5 | 22.2  | 3005     | 2 T08841            | polyprotein - dour                                |
| 13         | 342.5 | 18.6  | 2070     | 2 T08839            | polyprotein - dour                                |
| 14         | 102.5 | 5.6   | 660      | 1 VHHWH2            | structural protein                                |
| 15         | 102.5 | 5.6   | 692      | 2 H71426            | hypothetical protein                              |
| 16         | 101   | 5.6   | 564      | 2 S36637            | tetracycline 6-hydro                              |
| 17         | 101   | 5.5   | 600      | 2 B46642            | fatty-acid synthas                                |
| 18         | 100.5 | 5.5   | 353      | 2 G87392            | receptor adenylyl                                 |
| 19         | 99    | 5.4   | 399      | 2 AH3038            | CDA peptide synt                                  |
| 20         | 99    | 5.4   | 399      | 2 C92847            | transferrin precu                                 |
| 21         | 97.5  | 5.3   | 1085     | 2 T03531            | hypothetical protein                              |
| 22         | 95.5  | 5.2   | 470      | 2 JC4048            | signal recognitio                                 |
| 23         | 94.5  | 5.1   | 2795     | 2 JG4743            | DNA-directed DNA p                                |
| 24         | 93.5  | 5.1   | 1380     | 2 T18309            | conserved hypothet                                |
| 25         | 93    | 5.0   | 7463     | 2 T36248            | conserved hypothet                                |
| 26         | 92.5  | 5.0   | 706      | 2 S3371             | hypothetical protein                              |
| 27         | 92.5  | 5.0   | 715      | 2 G83612            | structural protein                                |
| 28         | 92    | 5.0   | 659      | 1 B44212            | genome polyprotein                                |
| 29         | 91.5  | 5.0   | 3414     | 1 GNWVC             |   |

#### ALIGNMENTS

RESULT 1

GNWVCJ

genome polyprotein - hepatitis C virus (strain J)

N; Contains capsid protein C; envelope protein M; major envelope protein NS4a; nonstructural protein NS4b; nonstructural protein NS5

C; Species: hepatitis C virus

C; Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 19-Jan-2001

C; Accession: A39253; P0086

R; Kato, N.; Hiiikata, M.; Ootsuyama, Y.; Nakagawa, M.; Ohkoshi, S.; Sugimura, T.; Shimoto, A; Title: Molecular cloning of the human hepatitis C virus genome from Japanese patients v

A; Reference number: A39253; MUID:91088550; PMID:2175903

A; Molecular type: genomic RNA

A; Residues: 1-3010 <NS4>

A; Cross-references: GB:092028; NID:9221610; PIDN:BA14233.1; PID:9221611

A; Title: Japanese isolates of the non-A, non-B hepatitis viral genome show sequence vari

A; Accession: P0086

A; Molecular type: genomic RNA

A; Residues: 2650-2707 <KA2>

A; Experimental source: Japanese isolate

C; Comment: The cleavage sites of this polyprotein have not been determined.

C; Superfamily: hepatitis C virus genome polyprotein

C; Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serine

F2-15/ Product: capsid protein C #status predicted <CPC>

F116-191/ Product: envelope protein M #status predicted <EM>

F19-389/ Product: major envelope protein E #status predicted <EM>

F390-729/ Product: nonstructural protein NS1 #status predicted <NS1>

F730-1006/ Product: nonstructural protein NS2 #status predicted <NS2>

F1007-1615/ Product: heparivirin #status predicted <NS3>

F1230-227/ Region: nucleotide-binding motif A (P-loop)

F1312-1317/ Region: nucleotide-binding motif B

F1316-1319/ Region: DEXH motif

F1616-1862/ Product: nonstructural protein NS4a #status predicted <NS4A>

F1863-2013/ Product: nonstructural protein NS4b #status predicted <NS4B>

F2014-3010/ Product: nonstructural protein NS5 #status predicted <NS5>

F196, 209, 234, 250, 305, 325, 417, 423, 430, 448, 532, 556, 576, 623, 645, 1213, 1255, 2041, 2077, 2240, 27

Query Match 96.3%; Score 1773; DB 1; Length 310;

Best Local Similarity 94.6%; Pred. No. 6.8e-141; Matches 333; Conservative 12; Mismatches 7; Indels 0; Gaps 0;

Ov 1 AHLQWVPPPLNVGRGDAILLTCAVHPPELFIDITKLLAIFGPMLVLOGITKVYPFR 60

Db 855 AHLQWVPPPLNVGRGDAILLTCAVHPPELFIDITKLLAIFGPMLVLOGITKVYPFR 914

Ov 61 AQLTRACMYRKAAGGHYQMAFNKLALTGTYYDHLPLQDWAGSLDVAEVPV 120

Db 915 AQLTRACMLVRKVAGHYQMAFNKLALTGTYYDHLPLQDWAGSLDVAEVPV 974

|  |  |  |  |
|--|--|--|--|
| QY   | 121 IFSPDMEVKLTWGA(DTAA)GGDIISGLPVSARRGRELILGPADNFEQQWRLLAPIYSQ 180  | Db   | 1095 GPITQMYTNVTDQDLVGWQAPGARSMTCTCGSDLYLVRHADVDFIPVRRGDSRGSLS 1154  |
| Db   | 975 VPSDMETKLTWTGADTAACGDIISGLPVSARRGRKELIIGPADSFGEQQRLLAPIYAQ 1034  | QY   | 301 PRPVSYLKGSGGPPLCPSPSCHA(VG)FRAAVCTRGVAKAVD(FIPV)ESMETMR 352  |
| QY   | 181 QTRGLLGCTTSLGRDKNOVEGEVQVSTATOSFLATCVCNGVWTVHAGSKTLGPK 240   | Db   | 1155 PRPVSYLKGSGGPPLCPSPSCHA(VG)FRAAVCTRGVAKAVD(FIPV)ESMETMR 1206  |
| Db   | 1035 QTRGLLGCTTSLGRDKNOVEGEVQVSTATOSFLATCVCNGVWTVHAGSKTLGPK 1094   | RESULT 3   | GNWVTW   |
| QY   | 241 GPITQMYTNVTDQDLVGWQAPGARSMTCTCGSDLYLVRHADVDFIPVRRGDSRGSLS 300  | QY   | Genome Polyprotein - hepatitis C virus (strain Taiwan)   |
| Db   | 1095 GPITQMYTNVTDQDLVGWQAPGARSMTCTCGSDLYLVRHADVDFIPVRRGDSRGSLS 1154  | C;Species: hepatitis C virus   | Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructural protein NS4a); nonstructural protein NS4b; nonstructural protein NS5       |
| QY   | 301 PRPVSYLKGSGGPPLCPSPSCHA(VG)FRAAVCTRGVAKAVD(FIPV)ESMETMR 352  | A;Note: Host Homo sapiens (man)  | Accession: A45573  |
| Db   | 1155 PRPVSYLKGSGGPPLCPSPSCHA(VG)FRAAVCTRGVAKAVD(FIPV)ESMETMR 1206  | C;Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001  | Accession: A40244  |
| RESULT 2   | A;Title: The Taiwanese hepatitis C virus genome: sequence determination and mapping the genome Polyprotein - hepatitis C virus (strain Taiwan)                   | R;Chen, P.J.; Lin, M.H.; Tai, K.F.; Liu, P.C.; Lin, C.J.; Chen, D.S.   | Title: The Taiwanese hepatitis C virus genome: sequence determination and mapping the genome Polyprotein - hepatitis C virus (strain Taiwan)                         |
| A;45573  | N;Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructural protein NS4a); nonstructural protein NS4b; nonstructural protein NS5 | R;Virology 188, 102-113, 1992  | Virology 188, 102-113, 1992  |
| A;Cross-references: GB:D11168; GB:D0171; NID:9221612; PID:9221613  | A;Experimental source: HCV-JT  | A;Reference number: A40244; MUID:9223206; PMID:1314449   | A;Reference number: A40244; MUID:9223206; PMID:1314449   |
| A;Sequence: sequence extracted from NCBI backbone (NCBIN:106206, NCBI:P:106207)  | A;Species: hepatitis C virus   | A;Molecule type: genomic RNA   | A;Molecule type: genomic RNA   |
| C;Superfamily: hepatitis C virus genome polyprotein  | C;Accession: A45573  | A;Residues: 1-301 <CHE>  | A;Residues: 1-301 <CHE>  |
| C;Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin   | A;Text: Molecular cloning of hepatitis C virus genome from a single Japanese carrier: S  | A;Cross-references: GB:084754  | A;Cross-references: GB:084754  |
| P;1-115/Product: capsid protein C #status predicted <PC>   | R;Virals Res. 23, 39-53, 1992  | C;Keywords: ATP; capsid protein; envelope protein; glyccoprotein; hydrolase; nonstructural protein NS1 #status predicted <PC>  | C;Keywords: ATP; capsid protein; envelope protein; glyccoprotein; hydrolase; nonstructural protein NS1 #status predicted <PC>  |
| P;1-16-191/Product: envelope protein M #status predicted <EP>  | R;Virus Res. 23, 39-53, 1992   | F;1-115/Product: capsid protein C #status predicted <CP>   | F;1-115/Product: capsid protein C #status predicted <CP>   |
| P;1-92-389/Product: major envelope protein E #status predicted <ME>  | R;Virus Res. 23, 39-53, 1992   | F;1-16-192/Product: envelope protein M #status predicted <EP>  | F;1-16-192/Product: envelope protein M #status predicted <EP>  |
| P;378-729/Product: nonstructural protein NS1 #status predicted <NS1>   | R;Virus Res. 23, 39-53, 1992   | F;1-19-389/Product: major envelope protein E #status predicted <ME>  | F;1-19-389/Product: major envelope protein E #status predicted <ME>  |
| P;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  | R;Virus Res. 23, 39-53, 1992   | F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>   | F;390-729/Product: nonstructural protein NS1 #status predicted <NS1>   |
| P;007-1615/Product: hepacivirin #status predicted <NS3>  | R;Virus Res. 23, 39-53, 1992   | F;1007-1615/Product: hepacivirin #status predicted <NS3>   | F;1007-1615/Product: hepacivirin #status predicted <NS3>   |
| P;1-1237/Region: nucleotide-binding motif A (P-loop)   | R;Virus Res. 23, 39-53, 1992   | F;120-127/Region: nucleotide-binding motif A (P-loop)  | F;120-127/Region: nucleotide-binding motif A (P-loop)  |
| P;1-12-1317/Region: nucleotide-binding motif B   | R;Virus Res. 23, 39-53, 1992   | F;1312-1317/Region: DEXH motif   | F;1312-1317/Region: DEXH motif   |
| P;1-16-1319/Region: DEXH motif   | R;Virus Res. 23, 39-53, 1992   | F;1616-1862/Product: nonstructural protein NS4a #status predicted <NA4>  | F;1616-1862/Product: nonstructural protein NS4a #status predicted <NA4>  |
| P;1-163-2013/Product: nonstructural protein NS5 #status predicted <NS5>  | R;Virus Res. 23, 39-53, 1992   | F;1863-2013/Product: nonstructural protein NS5 #status predicted <NS5>   | F;1863-2013/Product: nonstructural protein NS5 #status predicted <NS5>   |
| P;1-196-2137/Region: hepacivirin #status predicted <NS3>   | R;Virus Res. 23, 39-53, 1992   | F;2014-3010/Product: nonstructural protein NS2 #status predicted <NS2>   | F;2014-3010/Product: nonstructural protein NS2 #status predicted <NS2>   |
| P;1-112-1319/Region: nucleotide-binding motif B  | R;Virus Res. 23, 39-53, 1992   | F;196-209/233,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,2077,2127,2137/Region: nucleotide-binding motif B                               | F;196-209/233,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,2077,2127,2137/Region: nucleotide-binding motif B                               |
| P;1-16-1319/Region: DEXH motif   | R;Virus Res. 23, 39-53, 1992   | Query Match 95.0%; Score 1749; DB 1; Length 3010; Best Local Similarity 92.9%; Pred. No. 7.2e-139; Matches 327; Conservative 92.9%; MisMatches 14; Indels 0; Gaps 0; | Query Match 95.0%; Score 1749; DB 1; Length 3010; Best Local Similarity 92.9%; Pred. No. 7.2e-139; Matches 327; Conservative 92.9%; MisMatches 14; Indels 0; Gaps 0; |
| P;1-163-2013/Product: nonstructural protein NS4a #status predicted <NA4>   | R;Virus Res. 23, 39-53, 1992   | Matches 327; Conservative 92.9%; MisMatches 14; Indels 0; Gaps 0;  | Matches 327; Conservative 92.9%; MisMatches 14; Indels 0; Gaps 0;  |
| P;1-2014-3010/Product: nonstructural protein NS5 #status predicted <NS5>   | R;Virus Res. 23, 39-53, 1992   | QY 1 AHLQWIPNPNVRRGGRDAILTCAVHPPLIFDTIKLLAIFGLPMIQAQTGVYFV 60  | QY 1 AHLQWIPNPNVRRGGRDAILTCAVHPPLIFDTIKLLAIFGLPMIQAQTGVYFV 60  |
| Query Match 95.9%; Score 1766; DB 1; Length 3010; Best Local Similarity 94.9%; Pred. No. 2.7e-140; Matches 334; Conservative 97%; MisMatches 11; Indels 0; Gaps 0; | Matches 334; Conservative 97%; MisMatches 11; Indels 0; Gaps 0;  | Db 855 AHLQWIPNPNVRRGGRDAILTCAVHPPLIFDTIKLLAIFGLPMIQAQTGVYFV 914   | Db 855 AHLQWIPNPNVRRGGRDAILTCAVHPPLIFDTIKLLAIFGLPMIQAQTGVYFV 914   |
| QY 1 AHLQWIPNPNVRRGGRDAILTCAVHPPLIFDTIKLLAIFGLPMIQAQTGVYFV 60  | Db 855 AHLQWIPNPNVRRGGRDAILTCAVHPPLIFDTIKLLAIFGLPMIQAQTGVYFV 914   | QY 121 IFSPDMEVKLTWGA(DTAA)GGDIISGLPVSARRGRELILGPADNFEQQWRLLAPIYSQ 180   | QY 121 IFSPDMEVKLTWGA(DTAA)GGDIISGLPVSARRGRELILGPADNFEQQWRLLAPIYAQ 1034  |
| Db 855 AHLQWIPNPNVRRGGRDAILTCAVHPPLIFDTIKLLAIFGLPMIQAQTGVYFV 914   | Db 855 AHLQWIPNPNVRRGGRDAILTCAVHPPLIFDTIKLLAIFGLPMIQAQTGVYFV 914   | Db 975 VPSDMETKLTWTGADTAACGDIISGLPVSARRGRKELIIGPADSFGEQQRLLAPIYAQ 1034   | Db 975 VPSDMETKLTWTGADTAACGDIISGLPVSARRGRKELIIGPADSFGEQQRLLAPIYAQ 1034   |
| QY 61 AQLIRACMVKAGGHVQMAFKLALLTGTVYDHTPLODWAHGLRLLAVAPV 120  | QY 61 AQLIRACMVKAGGHVQMAFKLALLTGTVYDHTPLODWAHGLRLLAVAPV 120  | QY 181 QTRGLLGCTTSLGRDKNOVEGEVQVSTATOSFLATCVCNGVWTVHAGSKTLGPK 240  | QY 181 QTRGLLGCTTSLGRDKNOVEGEVQVSTATOSFLATCVCNGVWTVHAGSKTLGPK 240  |
| Db 915 AQLIRACMVKAGGHVQMAFKLALLTGTVYDHTPLODWAHGLRLLAVAPV 974   | Db 915 AQLIRACMVKAGGHVQMAFKLALLTGTVYDHTPLODWAHGLRLLAVAPV 974   | Db 1035 QTRGLLGCTTSLGRDKNOVEGEVQVSTATOSFLATCVCNGVWTVHAGSKTLGPK 1094  | Db 1035 QTRGLLGCTTSLGRDKNOVEGEVQVSTATOSFLATCVCNGVWTVHAGSKTLGPK 1094  |
| QY 121 IFSPDMEVKLTWGA(DTAA)GGDIISGLPVSARRGRELILGPADNFEQQWRLLAPIYSQ 180   | QY 121 IFSPDMEVKLTWGA(DTAA)GGDIISGLPVSARRGRELILGPADNFEQQWRLLAPIYAQ 1034  | QY 241 GPITQMYTNVTDQDLVGWQAPGARSMTCTCGSDLYLVRHADVDFIPVRRGDSRGSLS 300   | QY 241 GPITQMYTNVTDQDLVGWQAPGARSMTCTCGSDLYLVRHADVDFIPVRRGDSRGSLS 300   |
| Db 975 VPSDMETKLTWTGADTAACGDIISGLPVSARRGRKELIIGPADSFGEQQRLLAPIYAQ 1034   | Db 975 VPSDMETKLTWTGADTAACGDIISGLPVSARRGRKELIIGPADSFGEQQRLLAPIYAQ 1034   | Db 1095 GPITQMYTNVTDQDLVGWQAPGARSMTCTCGSDLYLVRHADVDFIPVRRGDSRGSLS 1154   | Db 1095 GPITQMYTNVTDQDLVGWQAPGARSMTCTCGSDLYLVRHADVDFIPVRRGDSRGSLS 1154   |
| QY 181 QTRGLLGCTTSLGRDKNOVEGEVQVSTATOSFLATCVCNGVWTVHAGSKTLGPK 240  | QY 181 QTRGLLGCTTSLGRDKNOVEGEVQVSTATOSFLATCVCNGVWTVHAGSKTLGPK 240  | Db 1155 PRPVSYLKGSGGPPLCPSPSCHA(VG)FRAAVCTRGVAKAVD(FIPV)ESMETMR 1206   | Db 1155 PRPVSYLKGSGGPPLCPSPSCHA(VG)FRAAVCTRGVAKAVD(FIPV)ESMETMR 1206   |
| Db 1035 QTRGLLGCTTSLGRDKNOVEGEVQVSTATOSFLATCVCNGVWTVHAGSKTLGPK 1094  | Db 1035 QTRGLLGCTTSLGRDKNOVEGEVQVSTATOSFLATCVCNGVWTVHAGSKTLGPK 1094  | QY genome polyprotein - hepatitis C virus (isolate JK1) S18030   | QY genome polyprotein - hepatitis C virus (isolate JK1) S18030   |
| QY 241 GPITQMYTNVTDQDLVGWQAPGARSMTCTCGSDLYLVRHADVDFIPVRRGDSRGSLS 300   | QY 241 GPITQMYTNVTDQDLVGWQAPGARSMTCTCGSDLYLVRHADVDFIPVRRGDSRGSLS 300   | RESULT 4   | RESULT 4   |

A: Contains: capsid protein C; envelope protein M; hepatitis C virus  
 C: Species: hepatitis C virus  
 C: Contains: nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
 A: Variety: isolate JKL  
 C: Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 23-Mar-2001  
 C: Accession: S18030; S33570; A48332; S18029  
 R: Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.  
 Submitted to the EMBL Data Library, September 1991  
 A: Description: A whole genome of hepatitis C virus cDNA was isolated from a single patient.  
 A: Reference number: S18028  
 A: Accession: S18030  
 A: Molecule type: genomic RNA  
 A: Residues: 1-3010 <HON>  
 A: Cross-references: EMBL:X61596; NID:959478; PIDN:CA43793.1; PID:959479  
 A: Experimental source: isolate JKL from an individual  
 A: Note: the authors translated the codon AGG for residue 43 as Pro, TCG for residue 320 as Trp, and TTC for residue 771 as Ser  
 A: Note: sequence extracted from NCBI backbone (NCBIN:121747, NCBI:121748)  
 C: Superfamily: hepatitis C virus genome polyprotein  
 C: Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin  
 F: 2-15/ Product: capsid protein C #status predicted <CPC>  
 F: 116-191/ Product: envelope protein M #status predicted <EPM>  
 F: 192-389/ Product: major envelope protein E #status predicted <MEB>  
 F: 390-729/ Product: nonstructural protein NS1 #status predicted <NS1>  
 F: 730-1006/ Product: nonstructural protein NS2 #status predicted <NS2>  
 F: 1007-1615/ Product: hepacivirin #status predicted <NS3>  
 F: 1312-1317/ Region: nucleotide binding motif A (P-loop)  
 F: 1616-1862/ Product: nonstructural protein NS4 #status predicted <NS4>  
 F: 1863-2013/ Product: nonstructural protein NS4b #status predicted <NS4b>  
 F: 196-209, 234, 250, 305, 417, 423, 448, 532, 540, 556, 576, 623, 645/ Binding <site: carbohydrate (As  
 Query Match  
 Best Local Similarity 93.2%; Score 171; DB 1; Length 3010;  
 Matches 325; Conservative 11; Mismatches 16; Indels 0; Gaps 0;  
 Qy 1 AHHQWQIWPINVGGRDAIILTCVAVPELIFDITKLLAIFGPMVHQAGITKVYFVR 60  
 Db 855 ACUQWVWPVPINVGGRDAIILTCVAVHSELIFDITKLLAIFGPLMVHQAGITKVYFVR 914  
 Qy 61 AQGLIRAGMLVRKAAGHIVQAFMKAALTGIVYVHILIQDMHAGLDAVNEPV 120  
 915 AQGURIRACMLVRKVAAGHIVQAFMKAALTGIVYVHILIQDMHAGLDAVNEPV 974  
 Db 121 IFSDMEVKITWGADTAACGDTISGLGPVSRARRGREIILGPADNPFEGQWRKAPITYSQ 180  
 975 VPSDMEVKITWGADTAACGDTISGLGPVSRARRGREIILGPADFGRQWRKAPITYSQ 1034  
 Qy 181 QTRGLLGCTTSITGDRDKNQVEGEVQVSTATOSFLATCNGVQCVTFHAGSKTIAQPK 240  
 Db 1035 QTRGLFGCIVTSITGDRDNQVEGEAQVSTATOSFLATCNGVQCVTFHAGSKTIAQPK 1094  
 Qy 241 GRPTQMYVNDDQDILVGMAPPGRASRMPCTCSDSDDLYTWRADVPVRRGDSRSILLS 300  
 Db 1095 GPINQMYVNQDQDILVGMAPPGRASRMPCTCSDSDDLYTWRADVPVRRGDSRSILLS 1154  
 Qy 301 PRVSYLQKGSSECPILCSGHAVG3TIAAVCVRGVAQGAVDPIPVRENETMR 352  
 Db 1155 PRVSYLQKGSSECPILCSGHAVG3TIAAVCVRGVAQGAVDPIPVRENETMR 1206

GNWTC  
 genome polyprotein - hepatitis C virus  
 Nt:Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
 C:Species: hepatitis C virus  
 C:Date: 31-Mar-1992 #sequence revision 31-Mar-1992 #text\_change 19-Jan-2001  
 C:Accession: A38465  
 J.Takamizawa, A.; Mori, C.; Fuke, I.; Manabe, S.; Murakami, S.; Fujita, J.; Onishi, E.; I. Virol., 65, 1105-1113, 1991  
 A:Title: Structure and organization of the hepatitis C virus genome isolated from human  
 A:Reference number: A38465; MUID:91140698; PMID:1847440  
 A:Accession: A38465  
 A:Molecule type: genomic RNA  
 A:residues: 1-3101 <TAK>  
 A:Cross-References: EMBL:MS335; NID:932970; PID:AAA72945.1; PID:932971  
 C:Supplements: hepatitis C virus genome polyprotein  
 C:Keywords: AMP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural protein; nonstructural protein C #status predicted <CPC>  
 F:2-115/Product: capsid protein C #status predicted <BPM>  
 F:116-191/Product: envelope protein M #status predicted <BPM>  
 F:192-279/Product: major envelope protein E #status predicted <ME>  
 F:290-799/Product: nonstructural protein NS1 #status predicted <NS1>  
 F:330-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
 F:1007-1616/Product: hepacivirin #status predicted <NS3>  
 F:1230-1237/Region: nucleotide-binding motif A (P-loop)  
 F:1312-1319/Region: nucleotide-binding motif B  
 F:1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>  
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>  
 F:196, 209, 234, 250, 305, 325, 417, 423, 430, 448, 532, 540, 555, 576, 623, 645, 1213, 1255, 2041, 2077, 224  
 Query Match 92.2%; Score 1699; DB 1; Length 3101;  
 Best local Similarity 91.5%; Pred. No. 1.2e-134; Matches 322; Conservative 12; Mismatches 18; Indels 0; Gaps 0;  
 Qy 1 A H T L O W I P P I N V T G G R D A I L T L C A V H P L I F D T K U L A I F G P M I N Q A G I T K Y P F R 60  
 Db 855 A D L H W I P P I N V T G G R D A I L T L C A V H P L I F D T K U L A I F G P M I N Q A G I T K Y P F R 914  
 Qy 61 A Q G L I T R A C H M V R K A G G H V Q M A F M K L A M I G T Y V D H I T P L Q D W A H A G R D L A W A V E P V 120  
 Db 915 A Q G L I T R A C H M V R K A G G H V Q M A F M K L A M I G T Y V D H I T P L Q D W A H A G R D L A W A V E P V 974  
 Qy 121 I F S D M E V K I T T W G A D T A C G D I T S G L P Y S A R G R G E H I L G P A D N F G Q G W R L A P I T A Y S Q 180  
 Db 975 V F S D M E T K I T T W G A D T A C G D I T S G L P Y S A R R G K E H I L G P A D S D E G R G U R L A P I T A Y S Q 1034  
 Qy 181 Q T G E L I G C I T S L G R D K N Q V E R V Q V S T A T O S F L A C V G Y C T W F G G A G S K I L A P K 240  
 Db 1035 Q T G E L I G C I T S L G R D K N Q V E R V Q V S T A T O S F L A C V G Y C T W F G G A G S K I L A P K 1094  
 Qy 241 G P I T Q M Y T N D Q D I L G W Q A P P G A R S M T P C T G G S D L Y L V T R H A D V P Y R R G S R G S L S 300  
 Db 1095 G P I T Q M Y T N D Q D I L G W Q P K P P G A R S L T P C T G G S D L Y L V T R H A D V P Y R R G S R G S L S 1154  
 Qy 301 P R P V S Y L K Q S S G G P L L C P S G H A V G I F R A V C T R G V A K A V D F P V S E M E T M R 352  
 Db 1155 P R P V S Y L K Q S S G G P L L C P F G H A V G I F R A V C T R G V A K A V D F P V S E M E T M R 1206

RESULT 6

GNWTC  
 genome polyprotein - hepatitis C virus (strain HCV-1)  
 Nt:Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5  
 C:Species: hepatitis C virus  
 C:Date: 30-Sep-1992 #sequence revision 30-Sep-1992 #text\_change 19-Jan-2001  
 C:Accession: A39166; PQ0403; PQ0404;  
 R;Choo, Q.L.; Richman, K.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.; Coit,  
 Proc. Natl. Acad. Sci. U.S.A. 89, 2451-2455, 1991  
 A:Title: Genetic organization and diversity of the hepatitis C virus  
 A:Reference number: A39166; MUID:91172826; PMID:1848704  
 A:Accession: A39166

A; Molecule type: mRNA  
A; Residues: 1-3011 <CHO>  
A; Cross-references: GR:M62321; NID:9329873; PID:AAA45676.1; PIP:9329874  
R; Chan, S.W.; McCormish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, P.I.  
J. Gen. Virol., 73, 1131-1141, 1992  
A; Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to e  
A; Reference number: PQ0393; MUID:92268871; PMID:1316939  
A; Accession: PQ0403  
A; Molecule type: genomic RNA  
A; Residues: 1577-1633 <CH2>  
A; Cross-references: DDBJ:D10128  
A; Experimental source: isolates B-b17  
A; Accession: QD0404  
A; Status: preliminary  
A; Molecule type: genomic RNA  
A; Residues: 1577-1633 <CH2>  
A; Experimental source: isolates B-b17  
A; Cross-references: DDBJ:D10128  
A; Experimental source: isolates B-b16  
A; Accession: QD0404  
A; Status: preliminary  
A; Molecule type: genomic RNA  
A; Residues: 1577-1633 <CH2>  
A; Experimental source: isolates C #status predicted <CC>  
C; Superfamily: hepatitis C virus genome polyprotein  
C; Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural  
F1-115/Product: capsid protein M #status predicted <EPM>  
F116-191/Product: envelope protein M #status predicted <MEB>  
F1-192-389/Product: major envelope protein E #status predicted <NS1>  
F1-390-729/Product: nonstructural protein NS1 #status predicted <NS2>  
F1-1007-1615/Product: nonstructural protein NS2 #status predicted <NS3>  
F1-116-3011/Product: nonstructural protein NS3 #status predicted <NS4>  
F1-132-1317/Region: nucleotide binding motif B  
F1-133-1319/Region: DXH motif  
F1-162-1862/Product: nonstructural protein NS4a #status predicted <N4A>  
F1-1893-2031/Product: nonstructural protein NS4b #status predicted <N4B>  
F1-2014-30111/Product: nonstructural protein NS5 #status predicted <NS5>  
F1-1916-209, 234, 305, 325, 417, 423, 430, 448, 476, 532, 540, 556, 576, 623, 645, 1213, 1255, 2041, 2077, 22  
Query Match Similarity 87.7%; Score 1616; DB 1; Length 3011;  
Best Local Similarity 84.1%; Pred. No. 1\_2e-17;  
Matches 296; Conservative 28; Mismatches 28; Indels 0; Gaps 0;  
Db 1 AHLQWIPPLNVRRGRDAILLLTCAVHBLFDITKULLAIFGQLMLQAGITKVYFVR 60  
Db 855 AQLHWIPPLNVRRGRDAILLLTCAVHBLFDITKULLAIFGQLMLQAGITKVYFVR 60  
Db 61 AQLHTRACMLVKRAGGHVYQAFMKLAAALTGTYYVYDHTPLQDWAHGLRDIAVAPV 120  
Db 9.15 VQGLIRCALARKRKGHVYQWMIKLGALITGIVYVNHLPDRDWAHNGLRLAVAPV 974  
Qy 121 IFSLMEVKITTWGADTAAGGDISGLPVSARSGREILINGGPADNFESOGWRLLAPIYSQ 180  
Db 975 VFSQNETKLTWTGADTAAGGDISGLPVSARSGREILINGGPADNFESOGWRLLAPIYSQ 180  
Qy 181 QTRGLIGCITSLSGRDKNOVEGVQVYSTATOFLATCUNGVCWTVFHGSKTKLAPK 240  
Db 1035 QTRGLIGCITSLSGRDKNOVEGVQVYSTATOFLATCUNGVCWTVFHGSKTKLAPK 1094  
Qy 241 GPITQMYNTWDQDLVGWQAPGARSMTCTCGSDLYIUTRHADVPIRRRDSRSILS 300  
Db 1095 GPVQMYNTWDQDLVGWQAPGARSMTCTCGSDLYIUTRHADVPIRRRDSRSILS 1154  
Qy 301 PRPVSYLKGSSSGGPLICPGHAYGIFRAVCTRGVAKADFPIVESMTMR 352  
Db 1155 PRPVSYLKGSSSGGPLICPGHAYGIFRAVCTRGVAKADFPIVESMTMR 1206  
**RESULT 7**  
S40770 genome polyprotein - hepatitis C virus  
N; Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructural protein NS4a); nonstructural protein NS4b; nonstructural protein NS5  
C; Species: hepatitis C virus  
A; Note: host Homo sapiens (man)  
C; Date: 31-Dec-1992 #text\_change 19-Jan-2001  
C; Accession: A36814; A41546  
R; Inchauspe, G.; Zobedee, S.; Lee, D.H.; Sugitani, M.; Nasoof, M.; Prince, A.M.  
C; Species: hepatitis C virus  
C; Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 19-Jan-2001  
A; Description: Genomic structure of the human prototype strain H of hepatitis C virus: c  
C; Accession: A36814  
A; Accession: A36814  
A; Molecule type: genomic RNA  
A; Residues: 1-3011 <INC>

A;Cross-references: GB:N67463; NID:9329737; PID:AA45534.1; PID:9329738  
 R;Inchouaspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.  
 Proc. Natl. Acad. Sci. U.S.A. 88, 1029-1029, 1991  
 A;Title: Genomic structure of the human prototype strain H of hepatitis C virus: comparison  
 A;Reference number: A41546; MUID:92052256; PMID:1658800  
 A;Contents: annotation  
 A;Note: neither amino acid nor nucleotide sequence is given  
 C;Superfamily: hepatitis C virus genome polyprotein  
 C;Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural  
 F1-15/Product: capsid protein C #status predicted <CPC>  
 F1-16-191/Product: envelope protein M #status predicted <BPM>  
 F1-192-381/Product: major envelope protein E #status predicted <MEB>  
 F1-390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
 F1-730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
 F1-1007-1615/Product: hepacivirin #status predicted <NS3>  
 F1-1230-1237/Region: nucleotide-binding motif A (P-loop)  
 F1-1312-1317/Region: nucleotide-binding motif B  
 F1-1616-1862/Product: nonstructural protein NS4a #status predicted <NA4>  
 F1-1864-2011/Product: nonstructural protein NS4b #status predicted <NB4>  
 F1-2014-3011/Product: nonstructural protein NS5 #status predicted <NS5>  
 F1-196-209, 234, 305, 325, 417, 423, 430, 448, 476, 532, 540, 556, 576, 623, 645, 1213, 1255, 2041, 2240, 23  
 Query Match 87.1%; Score 1605; DB 1; Length 3011;  
 Best Local Similarity 83.5%; Pred. No. 1e-126;  
 Matches 294; Conservatve 30; Mismatches 28; Indels 0; Gaps 0;  
 Qy 1 AHLQWNPPLNVRGGRDADILLTCAVHPLBLIDITKLLAIFGLPMLQAGITKVPFV 60  
 Db 855 AQLHWNPPLNVRGGRDADILLTCAVHPLBLIDITKLLAIFGLPMLQAGITKVPFV 914  
 Qy 61 AOGHTRACMLVRKGAGHHVQMAFMKLALTGTYYDHLTPDWAHGLRLLAVEPV 120  
 Db 915 VQGLRLICRALAKTAGLGHVQMAIKLGALTGTCVNHLLAPLRLWAINGRLULAVEPV 974  
 Qy 121 IFSDMEVKITWGADTAACGDIISGLPVSAARRGRTTLLPADNFPGQGRLLAPITAYQ 180  
 Db 975 VFSRMETKLITWGADTAACGDIISGLPVSAARRGRTTLLPADNFPGQGRLLAPITAYQ 1034  
 Qy 181 QTRGLLGCLTSLSLGRDKQVEQVQVSTAQSFLATCVNQYCWTFHAGSKTLGPK 240  
 Db 1035 QTRGLLGCLTSLSLGRDKQVEQVQVSTAQSFLATCVNQYCWTFHAGSKTLGPK 1094  
 Qy 241 GRITQMYNTVUDQDQIIVGQAPPGRSMTRCTCCSSDLVTRADIVPVRREDSRSILS 300  
 Db 1095 GPVHQYQMYNTVUDQDQIIVGWPAPGSRSLTCTGSSDLVTRADIVPVRREDSRSILS 1154  
 Qy 301 PRPVSYLKGSSGGPGLCPGSGHAGVIFRAVCTRGVAKADDFIPFESMETMR 352  
 Db 1155 PRPVSYLKGSSGGPGLCPGSGHAGVIFRAVCTRGVAKADDFIPFESMETMR 1206  
 RESULT 9  
 JC5620 genome polyprotein - hepatitis C virus (isolate EUM1480) (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5)  
 C;Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 19-Jan-2001  
 C;Accession: JC5620  
 R;Chamberlain, R.W.; Adams, N.J.; Taylor, L.A.; Simmonds, P.; Elliott, R.M.  
 Biochem. Biophys. Res. Commun. 236, 44-49, 1997  
 A;Title: The complete coding sequence of hepatitis C virus genotype 5a, the predominant  
 A;Reference number: JC5620; MUID:97366593; PMID:9223423  
 A;Accession: JC5620  
 A;Molecule type: RNA  
 A;Residues: 1-3014 <CIR>  
 A;Cross-references: GB:Y1184  
 A;Experimental source: genotype 5a, which predominates in South Africa  
 A;Note: the translation of the nucleotide sequence is not complete in this paper  
 C;Superfamily: hepatitis C virus genome polyprotein  
 C;Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polypeptid; serin  
 F;2-115/Product: capsid protein C #status predicted <CPC>  
 F;116-191/Product: envelope protein M #status predicted <BPM>  
 F;192-381/Product: major envelope protein E #status predicted <MEB>  
 F;384-408/Region: hypervariable #status predicted  
 F;390-730/Product: nonstructural protein NS1 #status predicted <NS1>  
 F;731-1007/Product: nonstructural protein NS2 #status predicted <NS2>  
 F;1008-1616/Product: hepacivirin #status predicted <NS3>  
 F;1230-1238/Region: nucleotide-binding motif A (P-loop)  
 F;1312-1318/Region: nucleotide-binding motif B  
 F;1317-1320/Region: DEXH motif  
 F;1617-1863/Product: nonstructural protein NS4a #status predicted <NA4>  
 F;1864-2014/Product: nonstructural protein NS4b #status predicted <NB4>  
 F;2210-2249/Region: interferon sensitivity determining #status predicted  
 Query Match 76.1%; Score 1401; DB 1; Length 3014;  
 Best Local Similarity 70.7%; Pred. No. 1.e-109;  
 Matches 249; Conservatve 51; Mismatches 52; Indels 0; Gaps 0;  
 Qy 1 AHLQWNPPLNVRGGRDADILLTCAVHPLBLIDITKLLAIFGLPMLQAGITKVPFV 60  
 Db 916 ABLARLICRLAKHLYVGKQYQALHLHRGTYVHAPKMDAASGLRTVATEP 975  
 Qy 121 IFSDMEVKITWGADTAACGDIISGLPVSAARRGRTTLLPADNFPGQGRLLAPITAYQ 180  
 Db 976 VFSAMETKIVTNGADTAACGDIISGLPVSAARRGRTTLLPADNFPGQGRLLAPITAYQ 1035  
 Db 916 QTRGLLGCLTSLSLGRDKQVEQVQVSTAQSFLATCVNQYCWTFHAGSKTLGPK 240  
 Db 1036 QTRGLLGCLTSLSLGRDKNEAEGEYOFISTATOTPLGICINGVWMTLFGAGSKTLGPK 1095  
 Qy 241 GRITQMYNTVUDQDQIIVGQAPPGRSMTRCTCCSSDLVTRADIVPVRREDSRSILS 300  
 Db 1096 GPVHQYQMYNTVUDQDQIIVGWPAPGSRSLTCTGSSDLVTRADIVPVRREDSRSILS 1155  
 Qy 301 PRPVSYLKGSSGGPGLCPGSGHAGVIFRAVCTRGVAKADDFIPFESMETMR 352  
 Db 1156 PRPVSYLKGSSGGPGLCPGSGHAGVIFRAVCTRGVAKADDFIPFESMETMR 1207  
 RESULT 10  
 GMWUJ genome polyprotein - hepatitis C virus (strain HC-J8) (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5)  
 C;Species: hepatitis C virus  
 C;Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 19-Jan-2001  
 C;Accession: AA40250; PQ0397; P00559  
 R;Okamoto, H.; Kurai, K.; Okada, S.I.; Yamamoto, K.; Lizuka, H.; Tanaka, T.; Fukuda, S.;  
 Virology 188, 331-341, 1992  
 A;Title: Full-length sequence of a hepatitis C virus genome having poor homology to reported  
 A;Reference number: A40250; MUID:92230232; PMID:131459  
 A;Accession: A40250  
 A;Molecule type: genomic RNA  
 A;Residues: 1-3033 <OKA>  
 A;Cross-references: GB:D10988; GB:D01221; NID:9221608; PID:BA01761.1; PID:9221609  
 R;Chan, S.W.; McOmisch, F.; Holmes, E.C.; Dow, B.; Peutherford, J.F.; Follett, E.; Yap, P.L.  
 J. Gen. Virol. 73, 1131-1141, 1992  
 A;Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to existing  
 A;Reference number: PQ0393; MUID:92268871; PMID:1316939  
 A;Accession: PQ0397  
 A;Molecule type: genomic RNA  
 A;Residues: 2678-2754 <CHA>  
 A;Cross-references: DDu:DI0134  
 A;Experimental source: isolate E-b12  
 R;Kato, N.; Ootsukawa, Y.; Okisaki, S.; Nakazawa, T.; Mori, S.; Hijikata, M.; Shimotohno,  
 Biochem. Biophys. Res. Commun. 181, 279-285, 1991  
 A;Title: Distribution of plural HCV types in Japan  
 A;Reference number: PQ0554; MUID:92068204; PMID:1720309  
 A;Accession: PQ0559

A; Molecule type: mRNA  
A; Residues: 2678-2729 <KAT>  
A; Cross-references: GB:D10562; GB:D00518; NID:9221523; PIDN:BAA01418.1; PID:g221524  
C; Superfamily: hepatitis C virus genome polyprotein  
C; Keywords: AMP; capsid; envelope protein; glycoprotein; hydrolase; nonstructural  
F; 1/15/2011/Product: capsid protein C #status predicted <CC>  
F; 116-191/Produced: envelope protein M #status predicted <EPM>  
F; 390-733/Product: major envelope protein E #status predicted <ME>  
F; 1934-1010/Produced: nonstructural protein NS1 #status predicted <NS1>  
F; 134-1619/Produced: hepacivirin #status predicted <NS3>  
F; 1234-1241/Region: nucleotide-binding motif A (P-loop)  
F; 1316-1321/Region: nucleotide binding motif B  
F; 1320-1323/Region: DEKH motif  
F; 1620-1866/Produced: nonstructural protein NS4a #status predicted <NA4>  
F; 1867-2017/Produced: nonstructural protein NS4b #status predicted <NA4B>  
F; 2018-3033/Produced: nonstructural protein NS5 #status predicted <NO5>  
F; 196,209,234,305,325,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,2038,23  
Query Match 70.7%; Score 1303; DB 1; Length 3033;  
Best Local Similarity 65.6%; Pred. No. 4,9e-10; Indels 0; Gaps 0;  
Matches 231; Conservative 56; Mismatches 67;  
Db 1 AHLQWQIPPLNVGRGDAILLTCAVHBLIPDITKLLIAIFCPLMVLQAGITKVYFVR 60  
Qy 859 AQIQWQVPLVERGGRGDLGIWAVILHRRIVFETVKOLLATIGPAVILKASLRITYFVR 918  
Db 61 AQLTRACMLVRAKGAGHVQAFMKLAAITGTGVYDHTPLQDWAHGLRDLAVALAEV 120  
Qy 919 AHALLRVCVTLVHLGARYIQMLLITGRWTGIVYDHLSPISTWAQGLDLAIVEV 978  
Db 121 IFSDMEVKITTWADTAAGDINGLPSVSRARRGEELINGLPAPNPEGGWRLLAPIYSQ 180  
Qy 979 VFSPEKKVIVWGAETVAGDHLGLPLVPSARLGREVELGPADGTSKWKLLAPIYQ 1038  
Db 181 QTRGLGICLTSLGDRDKNQEVSVQVSTATOSFLACNGVCWTVFHGGSKTLAQPK 240  
Qy 1039 CTGRGLGATVWSLTGRDKNEQAGOVQVLSVTOPFLGTSISCVLWTVHAGNKTLAGPK 1098  
Db 241 GPITOMTYNDQDUGWQAPPGARSMTCTCGSSDLYVTRHADIVURRDRGSRLS 300  
Qy 1099 GPVIVQMTSAGDVGWSPSPGKTSIDCTGAVDLYLVTNADVIPRKDRRGALLS 1158  
Db 301 PRPVVILKGSGGPILCPGSHAVIFRAVCTROVAKADFIVESMETTMR 352  
Qy 1159 PRPLSTIKSSGGPVLCSRGHAVIFRAAVCARGVAKSIDFIPVESUDVATR 1210  
Db 1159 PRPLSTIKSSGGPVLCPGSHAVIFRAAVCARGVAKSIDFIPVESUDVATR 1210

**RESULT 11**

JQ103 genome polyprotein - hepatitis C virus (isolate HC-J6) (nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5)  
C; Species: hepatitis C virus  
C; Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 17-Nov-2000  
C; Accession: JQ103  
R; Erter, J.C.; Desai, S.M.; Leary, T.P.; Chalmers, M.L.; Montes, C.C.; Mushahwar, I.K.  
J; Gen. Virol. 79, 41-45, 1998  
A; Title: Genomic analysis of two GB virus A variants isolated from captive monkeys.  
A; Reference number: Z16486; MUID:98120818; PMID:9460920  
A; Accession: T08841  
A; Status: translated from GB/EMBL/DDJB  
A; Molecule type: mRNA  
A; Residues: 1-3005 <ERK>  
A; Cross-references: EMBL:AF032425; NID:92828599; PIDN:AAC40502.1; PID:g2828600  
C; Superfamily: hepatitis C virus genome polyprotein  
C; Keywords: polyprotein

Query Match 22.2%; Score 408.5; DB 2; Length 3005;  
Best Local Similarity 31.3%; Pred. No. 1.1e-25; Indels 33; Gaps 9;  
Matches 111; Conservative 57; Mismatches 154; Indels 33; Gaps 9;

Qy 13 RGDRD---ATILLTCAVHBLIPDITKLLIAIFCPLMVLQAGITKVYFVRAGLIRAC 68  
Db 813 RGGDRKRVTAWWAGIIFREVRCASALTA-----LAALUDSDYLETI-ILATA 864  
Qy 69 MIVRK-----AGGYQAFMKLAAITGTGVYDHTPLQDWAHGLRDLAVALAEV 120  
Db 865 QPARAAFLDSLTFGLGDLTRAFTRRLRRGVLFQHQCSQSGAAATLXDLGVALEPV 924  
Qy 121 IFSDMEVKITTWADTAAGDINGLPSVSRARRGEELINGLPAPNPEGGWRLLAPIY 178  
Db 925 SVTARDLGICLTSLGDRDKNQEVSVQVSTATOSFLACNGVCWTVFHGGSKTLAQ 283  
Qy 179 SQCTRGLGICLTSLGDRDKNQEVSVQVSTATOSFLACNGVCWTVFHGGSKTLAQ 238  
Db 984 MORGUGFSVVKTMILGRDEREHEGSIIVLGTTSTRSMGIVCNGVMYTFHGSNARTAG 1043

QY 239 PKGIPITQMVNNVDQDLYGVHQAPPGARSMTPTCCTCASSDLVLTNTRHADVIFVRRGDSRSI 298  
Db :|||: ;|||: ;|||: ;|||: ;|||: ;|||: ;|||: ;|||: ;|||: ;|||: ;|||: ;|||: ;|||:  
1044 FVGPNCRWSPSPDVAWVPLPSASCERPCKGCTOTSWCIRN--DGALCHAPLSKVEL 1101

QY 299 LSPRPPSVLKGSSGGPLCPLCPSGHAGIFRAAVCTRGV-----AKAVDFIPVES 346  
Db :|||: ;|||: ;|||: ;|||: ;|||: ;|||: ;|||: ;|||: ;|||: ;|||: ;|||:  
1102 DLPTBEISDRGSSGSPILCDEGHVQGMV-VSULHRGVKVTGKVYVKPWTLPKDS 1155

RESULT 13

Polyprotein - marmoset hepatitis GB virus A

C;Species: marmoset hepatitis GB virus A

C;Date: 23-Jul-1999 #sequence\_revision 23-Jul-1999 #text\_change 17-Nov-2000

C;Accession: T08339 Rinker, J.C.; Detali, S.M.; Leary, T.P.; Chalmers, M.L.; Montes, C.C.; Mushahwar, I.K.

A;Title: Genomic analysis of two GB virus A variants isolated from captive monkeys.

A;Reference number: Z16486; MUID:98120818; PMID:9460920

A;Accession: T08339 A;Status: translated from GB/EMBL/DDBJ

A;Molecule type: genomic RNA

A;Residues: 1-2970 &lt;ERK&gt;

A;Cross-references: EMBL:AF023424; NID:9282857; PID:AC40501.1; PID:92828598

C;Supergroup: hepatitis C virus genome polyprotein

C;Keywords: structural protein

Query Match Best Local Similarity 5.6%; Score 102.5; DB 1; Length 690;

Matches 82; Conservative 52; Mismatches 130; Indels 151; Gaps 19;

Db 35 TXKLL--AIRGPIMVQAG-----ITKVYPFVR---AGSILRAMVLRKAAGCHYQMA 83

Db 151 TNELVLYKAPLSPLPLPQDGNTNHIMATEASNYAQTRVRAATIRYRPLPVNAVGGAIS 210

Db 84 FMKLAALTGTVVYDHITPLQWAHGLRDLAVAVPVIPIFSDMEVILITKGADTAACDII 143

Db 211 FWPOQTTPITPSV-----DMNSITSTDVYRLVQPGIASELVI----- 246

Db 144 SGLPVSAARRGRELIGPAD--NFEGGWRLAPI-TAYSQOTRGILL---GCITSLTG 195

Db 247 -----P-SERLHYRQWGSWSVETCSVABEATSGVLMCITHGSLSVSYTN 290

Db 196 -----RDKNQVEGEVQVSVTQSTQSL 216

Db 291 TPYTGALGILDAFALEEFRRUTPQGMINTNRVSRYSTARHLRQGADGMELTTAATRM 350

Db 217 A---TCVNGV-----CMVVFH-----GAG-----SNTLAGBKG-PIT 244

Db 351 KDLFYFISTNGVGEIGRGIATLFNLADTLIGGLPTELISAGGOLFYSRPSVVSANGEPIV 410

Db 245 QMYTNYNDQDLYGVHQAPPGRASMTPTCGSDSLVY--TRHADVIFVRRGDSRG-SLIS 300

Db 411 KLYTSENA----QDGKTAIPHIDLDLGSRVVIQDYDNOHEODPRTRPAPSRPFSYR 466

Db 467 ANDVWLSLTAAEVYDOSTYGSSTGPVVY--SDSVTLVNVATGQAVERSSLDWTKV 519

RESULT 14

VWWH2 structural protein 2 precursor - hepatitis E virus (strain Burma)

C;Species: hepatitis E virus

C;Date: 30-Sep-1992 #sequence\_revision 30-Sep-1992 #text\_change 23-Jul-1999

C;Accession: C00778 R;Tam, A.W.; Guerra, M.E.; Huang, C.C.; Bradley, D.W.; Fry, K.E.; Reyes, G.

Virology 185, 120-131, 1991 Hepatitis E virus (HEV): molecular cloning and sequencing of the full-length viral genome. Hepatitis E virus (HEV): molecular cloning and sequencing of the full-length viral genome. Hepatitis E virus (HEV): molecular cloning and sequencing of the full-length viral genome.

A;Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of *Arabidopsis thaliana*. A;Reference number: H71426; MUID:9812113; PMID:9461215

A;Accession: H71426 A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-692 &lt;BEV&gt;

A;Cross-references: GB:Z97340; NID:92244950; PID:e327492; PID:92244965

C;Genetics: A;Map position: 4COP9-4G3845

Query Match Best Local Similarity 5.6%; Score 102.5; DB 2; Length 692;

Matches 91; Conservative 41; Mismatches 127; Indels 121; Gaps 22;

Db 8 PPLNVRGERDAITLTCWHPPELFDITKLLAIFGLMLVLOGGITKPY--FVRQGLI 65

Db 101 PTASVYAGNCIEL---KGPEIMDLTELLR--FLTMML---FSKRPFAVFELESAGTY 151

Db 66 RACMLVR-KAAGHYWOMAF-----KIAALTGYVYVPHLT 101

A;Note: the authors translated the codon CGC for residue 2 as Ala

C;Superfamily: hepatitis E virus structural protein 2

QY 102 LDQWAHAGLRLDVAVAPVPIVSDFMEKLTWGAIDTACGDTISGLVNSARRGREILLGP 1611  
 Db 212 LHD--GGLSNUVLGY-----AHOG-----MVAAR----- 2344  
 QY 162 DNPEGQGN--RLLAIPITAYSOOTRGGLIGCITSLTRGDKNQVEGVQWVSTA---TOSF 2159  
 Db 235 -----WIAKUSVP-----CLIKAL---DEP-SPFQIQIHSLLGGGTASL 2700  
 QY 216 LATCVV----NGVCMVFEHGAGSKMILAGPKGP--ITOMYTNVDOLV--GWOAPPG 2633  
 Db 271 LTYLIREOKEFASATCTFATPGTPNMININGSGKHFTTING--SDLSPTFSAASSVDDL 3286  
 QY 264 RSMPTCIGSSDYLVLVRHADMV-TRRGDSRSGLSPPRPSVYKGSSGGPLCP--SG 3200  
 Db 329 RSEVTSSWSNDLDOVETRVLVSVYRSATAIGSRLPSIASAKAKVAGAGAILRPVSSG 3888  
 QY 321 HAVGIPRAAVCTRGVAGAVD 340  
 Db 389 TQPAFLNGC--GKIKCID 406

Search completed: May 6, 2004, 09:37:19  
Job time : 12.4639 secs





|                       |   |  |            |                        |             |
|-----------------------|---|--|------------|------------------------|-------------|
| FT                    | CARBOHYD  | 2240   | 2340       | N-LINKED (GLCNAC. . .) | (POTENTIAL) |
| FT                    | CARBOHYD  | 2529   | 2529       | N-LINKED (GLCNAC. . .) | (POTENTIAL) |
| FT                    | CARBOHYD  | 2788   | 2788       | N-LINKED (GLCNAC. . .) | (POTENTIAL) |
| SQ                    | SEQUENCE  | 3010 AA;   | 326573 MW; | 94A1C77435D642BB       | CRC64;      |
| Query Match           |   | 95.9%; Score 1766; DB 1; Length 3010;                                |            |                        |             |
| Best Local Similarity | 94.9%;  | Pred. No. 1. 2e-140;   |            |                        |             |
| Matches               | 334;  | Conservative 7; Mismatches 11;                                       | Indels 0;  | Gaps 0;                |             |
| QY                    |   | 1 AHLQWNPPLNRRGGRAAILLTCVHPELFDTIKULLAFGPMVHQAGITKTPVFPVR 60         |            |                        |             |
| Db                    |   | 855 AHLCWMPPLNVRGGDAAILLTCAHPELFDTIKULLALGPMVHQAGITKTPVFPVR 914      |            |                        |             |
| QY                    |   | 61 AQLIRACMLYRKAASCHYVQMAFKMKAALGTYVVDHLPQDWAHAGRLILAVAYEPV 120      |            |                        |             |
| Db                    |   | 915 AQLIRACMLYRKAASCHYVQMAFKMKAALGTYVVDHLPQDWAHAGRLILAVAYEPV 974     |            |                        |             |
| QY                    |   | 121 IFSDMVTKILWGAAPAAQGIIISGIPVSARGRGTEFLGPADNEFSGQRWLRALAPIYQS 180  |            |                        |             |
| Db                    |   | 975 VFSDMETKILWGAAPAAQGIIISGIPVSARGRGTEFLGPADNEFSGQRWLRALAPIYQS 1034 |            |                        |             |
| QY                    |   | 181 QTRGLIGCITSLSLGDKNQEGENQVQVSTATOSLATCVNGCWTYFHAGSKTLAGPK 240     |            |                        |             |
| Db                    |   | 1035 QTRGLIGCITSLSLGDKNQEGENQVQVSTATOSLATCVNGCWTYFHAGSKTLAGPK 1094   |            |                        |             |
| QY                    |   | 241 GPITQMTNTVQDLVWQAPPGARSIMPCTCGSSDXLVTRHADVLVPRRRGDSRSGILLS 300   |            |                        |             |
| Db                    |   | 1095 GPITQMTNTVQDLVWQAPPGARSIMPCTCGSSDXLVTRHADVLVPRRRGDSRSGILLS 1154 |            |                        |             |
| QY                    |   | 301 PRPVSYLKGSGGPPLCPSPGHAVGFLRAAFTGRVAKAVDIPVSEMETMR 352            |            |                        |             |
| Db                    |   | 1155 PRPVSYLKGSGGPPLCPSPGHAVGFLRAAFTGRVAKAVDIPVSEMETMR 1206          |            |                        |             |
| RESULT 3              |   |  |            |                        |             |
| POLG HCVTW            |   |  |            |                        |             |
| ID                    | POLG HCVTW  | STANDARD;  | PRT;       | 3010 AA.               |             |
| AC                    | P29416;   |  |            |                        |             |
| DT                    | 01-APR-1993   | (Rel. 25, Created)   |            |                        |             |
| DT                    | 01-APR-1993   | (Rel. 25, Last sequence update)                                      |            |                        |             |
| DT                    | 10-2003   | (Rel. 42, Last annotation update)                                    |            |                        |             |
| DE                    | Genome polyprotein [Contains: Capid protein C (core protein) (P22); Envelope glycoprotein E1 (GP2); Envelope glycoprotein E2 (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21) (EC 3.4.22.-); Protease/Helicase NS3 (P70) (Hepacivirus) (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)]. |  |            |                        |             |
| DE                    | Protein NS5A (P56); Nonstructural protein NS5B (P66) (RNA-directed RNA polymerase) (EC 2.7.7.48).   |  |            |                        |             |
| DE                    | NS4B (P27); Nonstructural protein NS4A (P4); Nonstructural protein NS5B (P66) (RNA-directed RNA polymerase) (EC 2.7.7.48).  |  |            |                        |             |
| OS                    | Hepatitis C Virus (isolate Taiwan) (HCV).   |  |            |                        |             |
| OC                    | Virions; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.  |  |            |                        |             |
| OX                    | NCBI_TAXID=31645;   |  |            |                        |             |
| RN                    | [1]   |  |            |                        |             |
| RP                    | SEQUENCE FROM N.A.  |  |            |                        |             |
| RX                    | MEDLINE:9223006; PubMed:1314449;  |  |            |                        |             |
| RA                    | Chen P.J., Lin M.H., Tai K.F., Liu P.C., Lin C.J., Chen D.S.;   |  |            |                        |             |
| RT                    | "The Taiwanese hepatitis C virus genome: sequence determination and mapping the 5' termini of viral genomic and antigenomic RNA.",  |  |            |                        |             |
| RL                    | Virology 188:102-113 (1992).  |  |            |                        |             |
| CC                    | -!- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are hydrophobic, suggesting a possible membrane-related function. NS3 and NS5 may play a role in the viral RNA replication.  |  |            |                        |             |
| CC                    | -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polyprotein, commonly with Asp or Glu in the P6 position, Cys or Tyr in P1 and Ser or Alanine in P1'.   |  |            |                        |             |
| CC                    | -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + (RNA) (N)   |  |            |                        |             |
| CC                    | -!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a lipoprotein envelope. The envelope consists of two proteins: protein M and glycoprotein E. The nucleocapsid is a complex of protein C and mRNA.  |  |            |                        |             |
| CC                    | -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.   |  |            |                        |             |

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; M84754; -; NOT\_ANNOTATED\_CDS.

CC PIR; A40244; GNWTW.

DR PDB; 1N64; 25-FEB-03.

DR DR PDB; INS3; 08-APR-98.

DR MEROPS; S29\_001; -.

DR M6ROPS; U39\_001; -.

DR InterPro; IPR00903; Cys\_Ser\_trypsin.

DR InterPro; IPR001410; DEAD.

DR InterPro; IPR02322; HCV\_capsid.

DR InterPro; IPR00251; HCV\_core.

DR InterPro; IPR00251; HCV\_env.

DR InterPro; IPR002331; HCV\_NS1.

DR InterPro; IPR002518; HCV\_NS2.

DR InterPro; IPR00745; HCV\_NS4a.

DR InterPro; IPR001190; HCV\_NS4b.

DR InterPro; IPR002368; HCV\_NS5a.

DR InterPro; IPR02166; HCV\_RdRp.

DR InterPro; IPR002519; HCV\_env.

DR InterPro; IPR002331; HCV\_NS1.

DR InterPro; IPR002518; HCV\_NS2.

DR InterPro; IPR00745; HCV\_NS4a.

DR InterPro; IPR001190; HCV\_NS4b.

DR InterPro; IPR002368; HCV\_capsid.

DR Pfam; PF01542; HCV\_core\_1.

DR Pfam; PF01539; HCV\_env\_1.

DR Pfam; PF01560; HCV\_NS1\_1.

DR Pfam; PF01538; HCV\_NS2\_1.

DR Pfam; PF02397; HCV\_NS3\_1.

DR Pfam; PF01061; HCV\_NS4a\_1.

DR Pfam; PF01001; HCV\_NS4b\_1.

DR Pfam; PF01506; HCV\_NS5a\_1.

DR Pfam; PF00271; helicase\_C\_1.

DR Pfam; PF00998; viral\_RdP\_1.

DR Pfam; PF0186062; HCV\_NS1\_1.

DR SMART; SM00487; DEXD\_C\_1.

KW Polypeptide; Transferase; RNA-directed RNA polymerase;

KW Core protein; Coat protein; Envelope protein; Helicase; AMP-binding;

KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;

KW 3D-structure.

FT INIT\_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE CELLULAR AMINOPEPTIDASE.

FT CHAIN 1 115 REMOVED FROM CAPSID PROTEIN C BY THE CORE PROTEIN (POTENTIAL).

FT CHAIN 115 191 MATRIX PROTEIN (POTENTIAL).

FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).

FT CHAIN 384 729 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).

FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).

FT CHAIN 107 1615 PROTEASE/Helicase NS3 (POTENTIAL).

FT CHAIN 1515 1862 NONSTRUCTURAL PROTEIN NS4 (POTENTIAL).

FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).

FT CHAIN 2014 3010 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).

FT TRANSMEM 347 369 CHARGE RELAY SYSTEM (BY SIMILARITY):

FT ACT\_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY):

FT ACT\_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY):

FT ACT\_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY):

FT NP\_35IND 1230 1237 ATP (POTENTIAL).

FT SITE 1316 1319 DECH BOX.

FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).

RESULT 4

POLG\_HCVBK ID POLG\_HCVBK STANDARD; PRT; 3010 AA.

POLG\_HCVBK ID P26663; DT 01-AUG-1992 (Rel. 23, Created) DT 01-AUG-1992 (Rel. 23, Last sequence update) DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Genome polypeptide (Contains: Capsid Protein C (Core protein) (P22); Envelope Glycoprotein (Contains: Envelope glycoprotein E1 (GP32); Envelope glycoprotein E2 (GP68)) (NS1); Protein P7; Nonstructural protein NS2 (P21)) (EC 3.4.22.-); Protease/helicase NS3 (P77) (Replicvirin) (EC 3.4.21.98.); Nonstructural protein NS4 (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P256); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)). DE NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48). DE Hepatitis C virus (isolate BK) (HCV). OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; OC Hepacivirus. OC NCBI\_TaxID:11105; RN [1]

SEQUENCE FROM N.A. RX MEDLINE:9110698; PubMed=1847440; RA Takamiawa A., Mori C., Fukuda T., Manabe S., Murakami S., Fujita J., RA Onishi B., Andoh T., Yoshida I., Okuyama H.; Structure and organization of the hepatitis C virus genome isolated RT from human carriers.; J. Virol. 65:1105-1113 (1991). RL [2]

SEQUENCE OF 1487-1500. RP MEDLINE:96235224; PubMed=8647104; RX MEDLINE:96235224; Borowski P., Heiland M., Oehlmann K., Becker B., Korneteky L.; RT "Non-structural protein 3 of hepatitis C virus inhibits

RT phosphorylation mediated by cAMP-dependent protein kinase.";  
 RL Bur. J. Biochem. 237:611-618(1996).  
 RN [3]  
 RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 1027-1215.  
 RX MEDLINE=9701598; PubMed=8861916;  
 RA Love R.A., Parge H.E., Wickerham J.A., Hostomsky Z., Habuka N.,  
 RA Moonaw E.W., Adachi T., Hostomská Z., Hostomsky Z.,  
 RT "The crystal structure of hepatitis C virus NS3 proteinase reveals a  
 RT trypsin-like fold and a structural zinc binding site.";  
 RL Cell 87:331-342(1996).  
 RN [4]  
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1027-1210 AND 1678-1691.  
 RX MEDLINE=98227846; PubMed=9568891;  
 RA Yan Y., Li Y., Munshi S., Sardana V., Cole J.L., Sardana M.,  
 RA Steinkeuer C., Tomei L., de Francesco R., Kuo L.C., Chen Z.;  
 RT "Complex of NS3 protease and NS4A peptide of BK strain hepatitis C  
 RL virus: a 2.2-A resolution structure in a hexagonal crystal form.";  
 CC Protein Sci. 7:837-847(1998).  
 CC -!- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are  
 CC hydrophobic, suggesting a possible membrane-related function. NS3  
 CC and NS5 may play a role in the viral RNA replication.  
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
 CC precursor polyprotein commonly with Asp or Glu in the P6  
 CC position, Cys or Thr in P1 and Ser or Ala in P1'.  
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate +  
 CC {RNA} (N)  
 CC -!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a  
 CC lipoprotein envelope. The envelope consists of two proteins:  
 CC protein M and glycoprotein E. The nucleocapsid is a complex of  
 CC protein C and mRNA.  
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

---

DR Pfam; PF01006; HCV\_NS4a; 1.  
 DR Pfam; PF01001; HCV\_Nsb; 1.  
 DR Pfam; PF01506; HCV\_Nsa; 1.  
 DR Pfam; PF0098; Viral\_RdRp; 1.  
 DR Prodrom; PD18602; HCV\_Ns1; 1.  
 DR SMART; SW00487; DExDC; 1.  
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA Polymerase;  
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;  
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;  
 KW 3D-structure.  
 FT INIT\_MET 1 1  
 FT INIT\_MET 1 1  
 FT CHAIN 1 115  
 FT CHAIN 116 191  
 FT CHAIN 192 383  
 FT CHAIN 384 729  
 FT CHAIN 730 1006  
 FT CHAIN 1007 1615  
 FT CHAIN 1616 1862  
 FT CHAIN 1863 2013  
 FT CHAIN 2014 3010  
 FT TRANSMEM 347 369  
 FT ACT SITE 1083 1083  
 FT ACT SITE 1107 1107  
 FT ACT SITE 1165 1165  
 FT NP\_BIND 1230 1237  
 FT SITE 1316 1319  
 FT CARBOHYD 195 196  
 FT CARBOHYD 209 209  
 FT CARBOHYD 234 234  
 FT CARBOHYD 250 250  
 FT CARBOHYD 305 305  
 FT CARBOHYD 417 417  
 FT CARBOHYD 423 423  
 FT CARBOHYD 430 430  
 FT CARBOHYD 448 448  
 FT CARBOHYD 532 532  
 FT CARBOHYD 540 540  
 FT CARBOHYD 556 556  
 FT CARBOHYD 576 576  
 FT CARBOHYD 623 623  
 FT CARBOHYD 645 645  
 FT CARBOHYD 2041 2041  
 FT CARBOHYD 2077 2077  
 FT CARBOHYD 2240 2240  
 FT CARBOHYD 2529 2529  
 FT CARBOHYD 2788 2788  
 FT STRAND 1031 1035  
 FT HELIX 1039 1047  
 FT STRAND 1050 1050  
 FT STRAND 1059 1063  
 FT STRAND 1068 1074  
 FT TURN 1075 1076  
 FT STRAND 1077 1081  
 FT HELIX 1082 1085  
 FT TURN 1086 1087  
 FT STRAND 1090 1092  
 FT TURN 1094 1094  
 FT STRAND 1095 1095  
 FT STRAND 1101 1103  
 FT TURN 1104 1107  
 FT STRAND 1108 1112  
 FT STRAND 1120 1120  
 FT STRAND 1122 1122  
 FT STRAND 1129 1133  
 FT TURN 1135 1136  
 FT STRAND 1140 1144  
 FT STRAND 1149 1157  
 FT HELIX 1158 1161  
 FT TURN 1162 1163  
 FT TURN 1165 1166  
 FT STRAND 1168 1171  
 FT TURN 1172 1174

FT STRAND 1175 1186  
 FT TURN 1187 1188  
 FT STRAND 1189 1197  
 FT HELIX 1198 1202  
 FT TURN 1203 1204  
 FT STRAND 1680 1688  
 SQ 3010 AA; 327189 MN; F8422D5ECCDFD9C CRC64;

Query Match 92.2%; Score 1699; DB 1; Length 3010;  
 Best Local Similarity 91.5%; Pred. No. 5.3e-135; Mismatches 322; Conservative 12; MisMatch 18; indels 0; Gaps 0;

OY 1 AHLQWVPPLNVRGGDAILLLTCVHPELJFDITKULLAIFGLMLVQAGITKUPFVR 60  
 855 ADLHWVPPLNVRGGDAILLLTCVHPELJFDITKULLAIFGLMLVQAGITKUPFVR 60  
 61 AGGLIRACMVLKRAAGHHVQMAFMKLAALGTTGTYVHDITPLQDWAHGAGRDLAVAEPV 120  
 915 AGGLIACMVLKRAAGHHVQMAFMKLAALGTTGTYVHDITPLQDWAHGAGRDLAVAEPV 974  
 OY 121 IFSDMEKTIWGAQDAAQGDTISGPVSAARRGRELGLGADNEFGQRLLPAPTA5Q 180  
 975 VPSDMETKTIWGAQDAAQGDTISGPVSAARRGRELGLGADNEFGQRLLPAPTA5Q 914  
 Db 181 QTRGLIGCTTSLTGDKNNQVEGEVNVVSTATOSFLATCNGVWTFHAGGSKTLAQPK 240  
 OY 1035 QPQLLSCCITSLTGDKNNQVEGEVNVVSTATOSFLATCNGVWTFHAGGSKTLAQPK 1094  
 Db 241 GITOMYTNPVDLIVQARPGARNTPCFCGSSILYLYRHALVYPERRGSRGSILS 300  
 1095 GPITOMYTNPVDLIVQARPGARNTPCFCGSSILYLYRHALVYPERRGSRGSILS 1154  
 Oy 301 PRPVSYVKGSSGPGLCPSCGAVGFRAACTGKAVDIFPEVEMTM MR 352  
 1155 PRPVSYVKGSSGPGLCPSCGAVGFRAACTGKAVDIFPEVEMTM MR 1206

RESULT 5  
 P01G HCV1 STANDARD: PRT: 3011 AA.  
 AC P2665; DT 01-AUG-1992 (Rel. 23, Created)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Genome PolyProtein [Contains: Capsid protein C (Core protein) (P22);  
 Envelope glycoprotein E1 (GP32); (GP35); Envelope glycoprotein E2  
 (GP68) (NS1); Protein p7; Nonstructural protein NS2 (P21)  
 (EC 3.4.22.-); Protease/Helicase NS3 (P70) (Hepatitis C virus)  
 (EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein  
 NS4B (P56); Nonstructural protein NS5A (P56); Nonstructural protein  
 NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].  
 OS Hepatitis C virus (isolate 1) (HCV).  
 OC Viruses; SSNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 NCBI\_TaxID:1104,  
 [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91172826; PubMed=1848704;  
 RA Choo Q.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C.,  
 RA Gallegos C., Coit D., Medina-Selby A., Barr P.J., Weiner A.J.,  
 RA Bradley D.W., Kuo G., Houghton M.;  
 RT "Genetic organization and diversity of the hepatitis C virus.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).  
 CC -!- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are  
 hydrophobic, suggesting a possible membrane-related function. NS3  
 and NS5 may play a role in the viral RNA replication.  
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
 precursor Polyprotein, commonly with Asp or Glu in the P5  
 position. Cys or Thr in P1 and Ser or Ala in P1'.  
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate +  
 (RNA) (N).  
 CC -!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a  
 CC lipoprotein envelope. The envelope consists of two proteins:

CC protein M and glycoprotein E. The nucleocapsid is a complex of  
 protein C and RNA.

CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to license@isb-sib.ch).

CC EMBL: M62321; AAC5676\_1;  
 DR PIR; A39166; GNWVC3.  
 DR PDB; 1A1V; 16-FEB-99.  
 DR PDB; 1HE; 25-NOV-98.  
 DR MEROPS; S29\_001; -.  
 DR IIPRO; IP009003; CY5 Ser\_trypsin.  
 DR IIPRO; IP01410; DEAD.  
 DR IIPRO; IP02521; HCV\_capsid.  
 DR IIPRO; IP002521; HCV\_core.  
 DR IIPRO; IP002519; HCV\_env.  
 DR IIPRO; IP002531; HCV\_NS1.  
 DR IIPRO; IP002518; HCV\_NS2.  
 DR IIPRO; IP00745; HCV\_NS4a.  
 DR IIPRO; IP001400; HCV\_NS4b.  
 DR IIPRO; IP002888; HCV\_NS5a.  
 DR IIPRO; IP002166; HCV\_RR9P.  
 DR IIPRO; IP001150; Helicase\_C.  
 DR IIPRO; IP004109; peptidase\_C29.  
 DR IIPRO; IP007035; RNA\_poli\_DS\_PS.  
 DR IIPRO; IP007094; RNA\_poli\_Pvir.  
 DR Pfam; PF01543; HCV\_capsid; 1.  
 DR Pfam; PF01542; HCV\_core; 1.  
 DR Pfam; PF01539; HCV\_env; 1.  
 DR Pfam; PF01560; HCV\_NS1; 1.  
 DR Pfam; PF01538; HCV\_NS2; 1.  
 DR Pfam; PF02907; HCV\_NS3; 1.  
 DR Pfam; PF01001; HCV\_NS4b; 1.  
 DR Pfam; PF01506; HCV\_NS5a; 1.  
 DR Pfam; PF00271; helicase\_C; 1.  
 DR Pfam; PF0098; viral\_RdRp; 1.  
 DR Pfam; PD186062; HCV\_NS1; 1.  
 DR SMART; SM0048; DDXDC; 1.  
 DR Polyprotein; Glycoprotein; Transferase; RNA-directed RNA-polymerase;  
 KW Core Protein; Coat Protein; Envelope Protein; Helicase; Amp-binding;  
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;  
 KW 3D-structure.  
 ET INIT\_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE  
 ET CHAIN 1 115 REMOVED FROM CAPSID PROTEIN C (POTENTIAL).  
 ET CHAIN 116 191 CAPSID PROTEIN C (POTENTIAL).  
 ET CHAIN 192 383 MATRIX PROTEIN (POTENTIAL).  
 ET CHAIN 384 729 MAJOR ENVELOPE PROTEIN E (POTENTIAL).  
 ET CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).  
 ET CHAIN 1007 1615 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).  
 ET CHAIN 1616 1862 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).  
 ET CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).  
 ET CHAIN 2014 3011 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).  
 ET TRANSMEM 347 369 POTENTIAL.  
 ET ACT\_SITE 1083 1083 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 ET ACT\_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 ET ACT\_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).  
 ET ATP (POTENTIAL).  
 ET NP\_BIND 1230 1237 ATP (POTENTIAL).  
 ET SITE 1316 1319 DECH BOX.  
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 417 423 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 476 476 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 540 540 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 556 556 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2077 2077 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2240 2240 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2364 2364 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 2789 2789 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ 3011 AA; 327197 NW; 65PFC9447FCBASF9 CRC64;

Query Match 87.7%; Score 1616; DB 1; Length 3011;  
 Best Local Similarity 84.1%; Pred. NC 5.5e-128;  
 Matches 296; Conservative 28; Mismatches 28; Indels 0; Gaps 0;

QY 1 AHIQWVWIPNVLGGGRDAITLICAVHBLFPIKULLAINGPLMTIQAGSTKVPFVR 60  
 855 AQLHWVWIPPNVRGERDAVILLMCAVHTLPFDITKLIAVFLPLWILQASILKVPFVR 914

QY 61 AQGILRACMLVKRAAGGHVQWMAFMKLAALGTQVYDHITPPQDWAGLRILAVATEP 120

QY ||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
 915 VQGLIRFCALAKRMIGGHVQWVQVTKLGALTGVYVNHLPDBWAINGLRLDAVAPV 974

QY 121 IFSDMEVKITIWGAATACGDDISGLPVSARGSRGSLILGPVQWLNQGCRPTIASKP 180  
 975 VFSQMETKLITWGADTAACGDLINGLPLVSARRGRELIGPAQDWVSKWRLAPITAYQ 1034

QY 181 OTRGILGCITSLGRDKQKVQEGVQVINSTASPLATCUNGSCTWTHGASSKTLAGPK 240

QY 1035 QTRGILGCITSLGRDKQKVQEGVQVINSTASPLATCUNGSCTWTHGASSKTLAGPK 1094

QY 241 GPTIOMYTNTDQDILVGWQAPPGARSMSMTPCTCGSSDLVYVTRHADVTIVPRRQDSRSLS 300

QY |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
 1095 GPVIQMWTQDQDVGWPAPQGSLSLPTCIGSSDLVYVTRHADVTIVPRRQDSRSLS 1154

QY 301 PRPVSYLKGSGGPLCPSGHAGVIFRAVCTGIVAKADFIVVESMMTTMR 352

QY 1155 PRPVSYLKGSGGPLCPAGHAVGIFRAVCTGIVAKADFIVVENLTMR 1206

RESULT 5

POLG\_HCVH STANDARD; PRT; 3011 AA.

ID P21958; AC T04155; DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Envelope glycoprotein [Contains: Capsid protein C (Core protein) (P22);  
 DE Glycoprotein E1 (GP32) (GP35); Envelope glycoprotein E2  
 (GP68) (GP70) (NS1); Protein F7; Nonstructural Protein NS2 (P21)  
 (EC 3.4.99.-); Protease/Helicase NS3 (P70) (Hepacivirus)  
 DE (EC 3.4.99.-); Nonstructural protein NS4A (P4); Nonstructural protein  
 NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein  
 NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].  
 DE Hepatitis C virus (isolate H) (HCV).  
 OC Viruses; srNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus.  
 OC NCBI TaxID:11108;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97331322; PubMed=9187654;  
 RA Yao N.; Hesson T.; Cable M.; Hong Z.; Kwong A.D.; Le H.V.; Weber P.C.;  
 RT "Structure of the hepatitis C virus RNA helicase domain.",  
 RL Nat. Struct. Biol. 4:463-467(1997).  
 RN [3]  
 RX X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.  
 RX MEDLINE=98154321; PubMed=9493277;  
 RA Kim J.I.; Morenstein K.A.; Griffith J.P.; Dwyer M.D.; Thomson J.A.;  
 RA Murcko M.A.; Lin C.; Caron P.R.;  
 RT "Hepatitis C virus NS3 RNA helicase domain with a bound oligonucleotide: the crystal structure provides insights into the mode of unwinding.",  
 RL Structure 6:89-100(1998).  
 CC -!- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.  
 CC -!- FUNCTION: NS3 IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN ESSENTIAL ROLE IN THE VIRUS REPLICATION.  
 CC -!- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE ACTIVATION OF NS3.  
 CC -!- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.  
 CC -!- FUNCTION: NS5B IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN ESSENTIAL ROLE IN THE VIRUS REPLICATION.  
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor Polyprotein, commonly with Asp or Glu in the P6 position, Cys or Thr in P1, and Ser or Ala in P1'.  
 CC -!- PROTEOLYTIC ACTIVITY: N nucleoside triphosphate + (RNA) (N).  
 CC -!- SIMILARITY: The NS3 protease belongs to peptidase family S39.  
 CC -!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a lipoprotein envelope. The envelope consists of two proteins: E1 and E2. The nucleocapsid is a complex of protein C and mRNA.  
 CC -!- PIM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY PROTEOLYTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.  
 CC -!- SIMILARITY: The NS2 protease belongs to peptidase family S29.  
 CC -!- This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation in the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement. (See <http://www.isb-sib.ch/announce/> or send an email to license@sb-sib.ch).

DR EMBL; M64643; AAA4534.1; -.

DR PIR; A36814; GNWCH.

DR PDB; 1HE1; 25-NOV-98.

DR PDB; 1AVV; 16-FEB-99.

DR PDB; 1A1R; 17-JUN-98.

DR MEROPS; S29\_001; -.

DR TRANSAC; T04155; -.

DR InterPro; IPR009003; Cys Ser\_trypsin.

DR InterPro; IPR001410; DEAF.

DR InterPro; IPR002522; HCV capsid.

DR InterPro; IPR002521; HCV core.

DR InterPro; IPR002519; HCV env.

DR InterPro; IPR002531; HCV\_N51.

DR InterPro; IPR002518; HCV\_N52.

DR InterPro; IPR007145; HCV\_N54.

DR InterPro; IPR001490; HCV\_N54b.

DR InterPro; IPR001490; HCV\_N54.

DR InterPro; IPR002868; HCV\_N5a.

DR InterPro; IPR00166; HCV\_RdRp.

DR InterPro; IPR001650; Helicase\_C.

DR InterPro; IPR00410; Peptidase\_C29.

DR InterPro; IPR007095; RNA\_Polymerase.

DR InterPro; IPR007094; RNA\_Polymerase.

DR Pfam; PF01543; HCV\_capsid; 1.

DR Pfam; PF01542; HCV\_core; 1.

DR Pfam; PF01539; HCV\_env; 1.

DR Pfam; PF01560; HCV\_N51; 1.

DR Pfam; PF01538; HCV\_N52; 1.

DR Pfam; PF02207; HCV\_N53; 1.

DR Pfam; PF01006; HCV\_N54; 1.

DR Pfam; PF01001; HCV\_N54b; 1.

|    |  |      |        |                                      |  |
|----|--|------|--------|--------------------------------------|--|
| DR | Pfam; PF01506; HCV_NSSa; 1.  | FT   | TURN   | 1419                                 | 1420   |
| DR | Pfam; PF00271; helicase_C; 1.  | FT   | STRAND | 1432                                 | 1436   |
| DR | Pfam; PF00998; viral_RdRp; 1.  | FT   | TURN   | 1438                                 | 1439   |
| DR | ProDom; P118602; HCV_NSS1; 1.  | FT   | STRAND | 1450                                 | 1453   |
| DR | SMART; SMD0487; DEXX1; 1.  | FT   | STRAND | 1456                                 | 1463   |
| KW | Protein; Glycoprotein; Transferase; RNA-directed RNA polymerase; Core protein; Coat protein; Envelope protein; Helicase; ATP-binding; Transmembrane; Nonstructural protein; Hydrolase; Serine protease; 3D-structure; 1. | FT   | STRAND | 1471                                 | 1478   |
| FT | INIT_MET   | 1    | 1      | 1                                    | 1  |
| FT | CHAIN  | 1    | 191    | REMOVED FROM CAPSID PROTEIN C BY THE |  |
| FT | CHAIN  | 192  | 383    | CELLULAR ANTIPEPTIDASE.              |  |
| FT | CHAIN  | 384  | 746    | CAPSID PROTEIN C.                    |  |
| FT | CHAIN  | 747  | 809    | ENVELOPE GLYCOPROTEIN E1.            |  |
| FT | CHAIN  | 810  | 1026   | ENVELOPE GLYCOPROTEIN E2.            |  |
| FT | CHAIN  | 1027 | 1557   | PROTEIN P7.                          |  |
| FT | CHAIN  | 1558 | 1711   | NONSTRUCTURAL PROTEIN NS2.           |  |
| FT | CHAIN  | 1712 | 1972   | PROTEASE/Helicase NS3.               |  |
| FT | CHAIN  | 1973 | 2420   | NONSTRUCTURAL PROTEIN NS4A.          |  |
| FT | TRANSMEM   | 2421 | 3011   | NONSTRUCTURAL PROTEIN NS4B.          |  |
| FT | ACT-SITE   | 3497 | 369    | POTENTIAL.                           |  |
| FT | ACT-SITE   | 1083 | 1083   | CHARGE RELAY SYSTEM (BY SIMILARITY). |  |
| FT | ACT-SITE   | 1107 | 1107   | CHARGE RELAY SYSTEM (BY SIMILARITY). |  |
| FT | NP-BIND  | 1165 | 1165   | CHARGE RELAY SYSTEM (BY SIMILARITY). |  |
| FT | SITE   | 1230 | 1237   | ATP (POTENTIAL).                     |  |
| FT | CARBHYD  | 1316 | 1319   | DECH BOX.                            |  |
| FT | CARBHYD  | 196  | 196    | ATP (POTENTIAL).                     |  |
| FT | CARBHYD  | 209  | 209    | N-LINKED (GLCNAC, .)                 | (POTENTIAL).   |
| FT | CARBHYD  | 234  | 234    | N-LINKED (GLCNAC, .)                 | (POTENTIAL).   |
| FT | CARBHYD  | 305  | 305    | N-LINKED (GLCNAC, .)                 | (POTENTIAL).   |
| FT | CARBHYD  | 417  | 417    | N-LINKED (GLCNAC, .)                 | (POTENTIAL).   |
| FT | CARBHYD  | 423  | 423    | N-LINKED (GLCNAC, .)                 | (POTENTIAL).   |
| FT | CARBHYD  | 430  | 430    | N-LINKED (GLCNAC, .)                 | (POTENTIAL).   |
| FT | CARBHYD  | 448  | 448    | N-LINKED (GLCNAC, .)                 | (POTENTIAL).   |
| FT | CARBHYD  | 476  | 476    | N-LINKED (GLCNAC, .)                 | (POTENTIAL).   |
| FT | CARBHYD  | 532  | 532    | N-LINKED (GLCNAC, .)                 | (POTENTIAL).   |
| FT | CARBHYD  | 540  | 540    | N-LINKED (GLCNAC, .)                 | (POTENTIAL).   |
| FT | CARBHYD  | 556  | 556    | N-LINKED (GLCNAC, .)                 | (POTENTIAL).   |
| FT | CARBHYD  | 576  | 576    | N-LINKED (GLCNAC, .)                 | (POTENTIAL).   |
| FT | CARBHYD  | 623  | 623    | N-LINKED (GLCNAC, .)                 | (POTENTIAL).   |
| FT | CARBHYD  | 645  | 645    | N-LINKED (GLCNAC, .)                 | (POTENTIAL).   |
| FT | TURN   | 1224 | 1226   | Db                                   | 975 VFSMETKLTWADTAAGDINGLPSARQOEELGPDAMVSKGRLRLLATYQ 1034  |
| FT | TURN   | 1232 | 1233   | Db                                   | 975 VFSMETKLTWADTAAGDINGLPSARQOEELGPDAMVSKGRLRLLATYQ 1034  |
| FT | TURN   | 1236 | 1238   | Qy                                   | 181 QTRGLIGCITSLTSGRDKNQVEEVSVQVSTATOSFLATCVNGCWTWTHGAGSKTLAGPK 240  |
| FT | HELIX  | 1239 | 1246   | Db                                   | 915 VQGLIRICALARKTAGGHVYOMAATKLGALTGTCVNHAPLDRWHDNGLDIAVEPY 974  |
| FT | TURN   | 1247 | 1248   | Qy                                   | 121 IFSPMEVKITIWAQDTAAQGDISGGLPVASRARREILEGPADFEGQGMRLLAPITAYQ 160   |
| FT | STRAND   | 1251 | 1255   | Db                                   | 1035 QTRGLIGCITSLTSGRDKNQVEEVSVQVSTATOSFLATCVNGCWTWTHGAGRTIASPK 1094   |
| FT | HELIX  | 1258 | 1271   | Qy                                   | 241 GPICOMTYNDQDVLGVWAPPGRSRSMTCCTCGSDLYVTRHADVIPRRGSRSILS 300   |
| FT | TURN   | 1272 | 1272   | Db                                   | 1095 GPVQIQTYYNDQDVLGVWPAPQGQSRSLTPCTCGSSDLYVTRHADVIPRRGSRSILS 1154  |
| FT | STRAND   | 1277 | 1280   | Qy                                   | 301 PRPSYLYKQSSGEGLCLCSGHSQHVAGIFRAVCTGQAKAVDFIVESMEMENTMR 352   |
| FT | STRAND   | 1283 | 1285   | Db                                   | 1155 PRPSYLYKQSSGEGLCLCPYQHGAGLFRAVCTGQAKAVDFIVESMEMENTMR 1206   |
| FT | HELIX  | 1296 | 1301   |                                      |  |
| FT | TURN   | 1302 | 1303   |                                      |  |
| FT | STRAND   | 1312 | 1316   |                                      |  |
| FT | TURN   | 1316 | 1319   |                                      |  |
| FT | HELIX  | 1323 | 1335   | RESULT 7                             |  |
| FT | STRAND   | 1336 | 1340   | POLG_HCVJ8                           |  |
| FT | STRAND   | 1343 | 1347   | ID_POLG_HCVJ8                        |  |
| FT | TURN   | 1352 | 1353   | STANDARD;                            |  |
| FT | TURN   | 1360 | 1361   | PRT;                                 | 3033 AA.   |
| FT | STRAND   | 1362 | 1366   | DT                                   | 01-AUG-1992 (Rel. 23, Last sequence update)  |
| FT | STRAND   | 1368 | 1368   | DT                                   | 28-FEB-2003 (Rel. 41, last annotation update)  |
| FT | STRAND   | 1373 | 1375   | DB                                   | Genome polypeptide (Contains: Capsid Protein C (Core protein) (P22); Envelope glycoprotein E1 (GP32); Envelope Glycoprotein E2 (GP68); GP70) (NS1); Protein P7; Nonstructural Protein NS2 (P21))   |
| FT | TURN   | 1376 | 1377   | DB                                   | (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirus)   |
| FT | STRAND   | 1378 | 1380   | DB                                   | (EC 3.4.21.-); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural Protein NS5A (P56); Nonstructural protein NS5B (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)). |
| FT | HELIX  | 1385 | 1389   | DB                                   | Hepatitis C virus (isolate HC-J5) (HCV).   |
| FT | STRAND   | 1389 | 1393   | OS                                   | Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  |
| FT | TURN   | 1397 | 1409   | OC                                   |  |
| FT | HELIX  | 1410 | 1411   |                                      |  |
| FT | STRAND   | 1414 | 1417   |                                      |  |



|          |   |
|----------|---|
| KW       | Transmembrane; Nonstructural protein; Hydrolase; Serine protease. |
| INIT-MET | REMOVED FROM CAPSID PROTEIN C BY THE CELLULAR AMINOPEPIDASE.      |
| FT       | CAPSID PROTEIN C (POTENTIAL).                                     |
| FT       | MATRIX PROTEIN (POTENTIAL).                                       |
| FT       | MAJOR ENVELOPE PROTEIN E (POTENTIAL).                             |
| FT       | NONSTRUCTURAL PROTEIN NS1 (POTENTIAL).                            |
| FT       | NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).                            |
| FT       | PROTEASE/HELICASE NS3 (POTENTIAL).                                |
| FT       | NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).                           |
| FT       | NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).                           |
| FT       | RNA-DIRECTED RNA POLYMERASE (POTENTIAL).                          |
| FT       | POTENTIAL.  |
| FT       | CHARGE RELAY SYSTEM (BY SIMILARITY).                              |
| FT       | CHARGE RELAY SYSTEM (BY SIMILARITY).                              |
| FT       | CHARGE RELAY SYSTEM (BY SIMILARITY).                              |
| FT       | ATP (POTENTIAL).  |
| FT       | DSCB BOX.   |
| FT       | N-LINKED (GLCNAC . .) (POTENTIAL).                                |
| FT       | N-LINKED (GLCNAC . .) (POTENTIAL).                                |
| FT       | N-LINKED (GLCNAC . .) (POTENTIAL).                                |
| FT       | N-LINKED (GLCNAC . .) (POTENTIAL).                                |
| FT       | N-LINKED (GLCNAC . .) (POTENTIAL).                                |
| FT       | N-LINKED (GLCNAC . .) (POTENTIAL).                                |
| FT       | N-LINKED (GLCNAC . .) (POTENTIAL).                                |
| FT       | N-LINKED (GLCNAC . .) (POTENTIAL).                                |
| FT       | N-LINKED (GLCNAC . .) (POTENTIAL).                                |
| FT       | N-LINKED (GLCNAC . .) (POTENTIAL).                                |
| FT       | N-LINKED (GLCNAC . .) (POTENTIAL).                                |
| FT       | N-LINKED (GLCNAC . .) (POTENTIAL).                                |
| FT       | N-LINKED (GLCNAC . .) (POTENTIAL).                                |
| FT       | N-LINKED (GLCNAC . .) (POTENTIAL).                                |
| FT       | N-LINKED (GLCNAC . .) (POTENTIAL).                                |
| FT       | N-LINKED (GLCNAC . .) (POTENTIAL).                                |
| FT       | N-LINKED (GLCNAC . .) (POTENTIAL).                                |
| FT       | N-LINKED (GLCNAC . .) (POTENTIAL).                                |
| FT       | N-LINKED (GLCNAC . .) (POTENTIAL).                                |
| FT       | N-LINKED (GLCNAC . .) (POTENTIAL).                                |
| FT       | N-LINKED (GLCNAC . .) (POTENTIAL).                                |
| FT       | N-LINKED (GLCNAC . .) (POTENTIAL).                                |
| FT       | N-LINKED (GLCNAC . .) (POTENTIAL).                                |
| FT       | N-LINKED (GLCNAC . .) (POTENTIAL).                                |
| FT       | N-LINKED (GLCNAC . .) (POTENTIAL).                                |
| FT       | N-LINKED (GLCNAC . .) (POTENTIAL).                                |
| FT       | N-LINKED (GLCNAC . .) (POTENTIAL).                                |
| SEQUENCE | 3033 AA; 329165 MW; F979F5CL2A73BEB9E CRC64;                      |

RESULT 10  
VST2 HEVPA  
ID - VST2 HEVPA  
AC P33426;  
STANDARD;  
PRT; 660 AA.

RESULT 11  
SR5C\_ARATH  
ID \_SR5C\_ARATH  
AC P37107; Q82570;  
STANDARD; PRT; 564 AA

DT 01-OCT-1994 (Rel. 30, Created)

DT 01-OCT-1994 (Rel. 30, last sequence update)

DT 15-MAR-2004 (Rel. 43, last annotation update)

DE Signal recognition particle 54 kDa protein, chloroplast precursor (SRP54) (54CP) (FPC).

DE FFC OR ATSGO940 OR P86150

OS Arabidopsis thaliana (Mouse-ear cress).

OC Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

OX NCBI\_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=cv. Columbia;

RX MEDLINE=9401817; PubMed=8408079;

RA Franklin A.E.; Hoffman N.E.;

RT "Characterization of a chloroplast homologue of the 54-kDa subunit of the signal recognition particle";

RT RL J. Biol. Chem. 268:22175-22180(1993).

[2]

RN SEQUENCE FROM N.A.

RC Amin P., Sy D., Pilgrim M., Barry D.H., Hoffman N.E.;

RT "Isolation of two Arabidopsis mutants in the nuclear gene ffc, encoding the 54 kDa subunit of chloroplast signal recognition particle";

RT Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.

[3]

RN SEQUENCE FROM N.A.

RC STRAIN=cv. Columbia;

RX MEDLINE=21016721; PubMed=11130714;

RA Tabata S., Kaneko T., Nakamura Y., Kotani H., Kato T., Asamizu E., Miyajima N., Sasamoto S., Kimura T., Hosouchi T., Kawashima K., Kohara M., Matsumoto M., Matsuno A., Muraki A., Nakayama S., Nakazaki N., Naruo K., Okumura T., Shinjo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M., Sato S., de la Bastide M., Huang E., Spiegel L., Gnoj L., O'Shaughnessy A., Preston R., Habermann K., Murray J., Johnson D., Ronking T., Nelson J., Nelson J., Stoenkemper T., Pepin K., Spiech J., Sekhon M., Armstrong J., Becker M., Beitter H., Cordum H., Cordes M., Courtney L., Courtney W., Dante M., Du H., Edwards J., Fryman J., Haakensen B., Lamar B., Latrelle P., Leonardi S., Meyer R., Mulvaney E., Oursky P., Riley A., Strowmatt C., Wagner-McPherson C., Willam A., Yoakum M., Bell M., Dedia N., Parnell L., Shah R., Rodriguez M., Hoorn See L., Vil D., Baker J., Kirchoff K., Roth K., King D., Bahret A., Miller B., Marra M.A., Martienssen R., McCombie W.R., Wilson R.K., Murphy G., Bancroft I., Volckaert G., Wambutt R., Duesterhoeft A., Stiekema W., Pohl T., Enlian K.-D., Terry N., Hartley N., Bent E., Johnson S., Langham S.-A., McCullagh B., Robben J., Grymonpre B., Zimmermann W., Ransperger U., Weiler H., Balke K., Wedler E., Peters S., Van Staeren M., Dirks W., Moeljiman P., Klein Lankhorst R., Weitzengger T., Botha G., Rose M., Hauf J., Bernreiser S., Hampel S., Ra Bents O., Lemcke K., Kolesov G., Mayer K., Rudd S., Schott H., Schueler C., Zaccaria P., Mewes H.-W., Bevan M., Fransz P., RT "Sequence and analysis of chromosome 5 of the plant Arabidopsis thaliana";

RT Nature 408:823-826 (2000).

RL -!- FUNCTION: May target chloroplast proteins to either the thylakoid or envelope membranes.

CC -!- SUBCELLULAR LOCATION: Chloroplast stroma.

CC -!- TISSUE SPECIFICITY: Most abundant in green shoot tissue and lower levels seen in the roots and isolated buds.

CC -!- SIMILARITY: Belongs to the GTP-binding SRP family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; AF092168; AAC64139.1; -.

DR EMBL; ALI62873; CAB85514.1; -.

DR PIR; S36037; S36637.

DR HSSP; O07347; IFPH.

DR InterPro; IPR003593; AAA ATPase.

DR InterPro; IPR000897; SRP54\_SPB.

DR InterPro; IPR004125; SRP54\_SPB.

DR InterPro; IPR004780; SRP\_sub.

DR Pfam; PF00448; SRP54; 1.

DR Pfam; PF0281; SRP54\_N\_1.

DR Pfam; PF02978; SRP\_SPB; 1.

DR ProDom; P000819; SRP54; 1.

DR SMART; SM0382; AAA; 1.

DR TIGRFAMs; TIGR0959; ffh; 1.

DR PROSITE; PS0000; SRP54; 1.

KW Signal recognition particle; GTP-binding; RNA-binding; Chloroplast; TRANSIT peptide.

FT CHAIN 76 564 SIGNAL RECOGNITION PARTICLE 54 kDa

FT DOMAIN 76 370 CHLOROPLAST.

FT DOMAIN 371 564 SIGNAL RECOGNITION PARTICLE 54 kDa

FT NP\_BIND 183 190 PROTEIN.

FT NP\_BIND 265 269 G-DOMAIN.

FT NP\_BIND 323 326 M-DOMAIN.

FT CONFLECT 76 75 G-TOP (BY SIMILARITY).

FT CONFLECT 76 75 GTP (BY SIMILARITY).

SO 564 AA; 61232 MW; 423F7285FB9063E4 CRC64;

Query Match 5 %; Score 101; DB 1; Length 564;

Best Local Similarity 26.1%; Pred. No. 0.65; Gaps 11;

Matches 54; Conservative 37; Mismatches 74; Indels 42; Gaps 11;

QY 26 VHPSELU---FDITKLALIFGPMLVLDQAGI-----TKQPFVRAQGLTRACMUVR 72

Db 154 VHDENKLMLGEVSLEFLPEAKSGPTVILLAGILOSVGKTTWCACYLKKHQG-KSCMLII 210

QY 73 KAAGHHIVQVQAFMKLAAL---TGTIVDHTPLO--DWAHAGIRDLAVALAEPVIFSDMEW 127

Db 211 --AGDVYRPAAIDLVILBQGVGPVYTAGTDVKPADIAKQGLKEAK-----NNNDV 261

QY 128 KITWGADTAACGDTISGIVVSASRGREIL---LGPAIDNEFGQGWRLLAPITAYSQQT 182

Db 262 VIM---DTAGRLQIDKGMDDELKDVKELNPTEVLLVVDAMQQ-BAAALVTFNWEI 315

QY 183 RGLUCGCTTSLTGDKRNQTEGEVONV 209

Db 316 -GITGAIITKLGDGSRGAGLSKEV 341

RESULT 12

DPO2\_MOUSE STANDARD; PRT; 600 AA.

AC P3611;

DT 01-FEB-1994 (Rel. 28, Created)

DT 01-FEB-1994 (Rel. 28, last sequence update)

DT 30-MAY-2000 (Rel. 39, last annotation update)

DN DNA Polymerase alpha 70 kDa subunit (DNA polymerase subunit B).

DN POLA2.

OS Mus musculus (Mouse).

CC Mus musculus (Mouse); Chordata; Craniata; Vertebrata; Buteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

CC NCBITaxID=10090;

RN [1]

RN SEQUENCE FROM N.A., AND SEQUENCE OF 84-102; 269-285 AND 394-403.

RX MEDLINE=93216788; PubMed=8463324;

RA Miyazawa H., Izumi M., Tada S., Takada R., Masutani M., Uti M., Hanoka F.;

RA "Molecular cloning of the cDNAs for the four subunits of mouse DNA polymerase alpha-prime complex and their gene expression during cell proliferation and the cell cycle.";

RA J. Biol. Chem. 268:8111-8122(1993).

CC -!- FUNCTION: May play an essential role at the early stage of chromosomal DNA replication by coupling the polymerase

CC EMBL; 221970; CAA79981.1; -.



|   |           |  |           |                          |         |
|---|-----------|--|-----------|--------------------------|---------|
|   | RESULT 14 | Db   | 290       | GNDVSPTFSPDGSKNAFVSTRGQP | 314     |
| TOLB_CHLIE  | ID        | TOLB CHLIE   | STANDARD; | PRT;                     | 434 AA. |
| AC  | ID        | TOLB CHLIE   | STANDARD; | PRT;                     | 434 AA. |
| Q8KEO0;   | AC        | TOLB CHLIE   | STANDARD; | PRT;                     | 434 AA. |
| 10-OCT-2003   | DT        | (Rel. 42, Created)   |           |                          |         |
| (Rel. 42, Last sequence update)   | DT        | 10-OCT-2003 (Rel. 42, Last annotation update)                |           |                          |         |
| TOLB protein precursor.   | DE        | TOLB OR CR06365.   |           |                          |         |
| Chlorobium tepidum.   | OS        | Bacterium; Chlorobi; Chlorobia; Chlorobiales; Chlorobiaceae; |           |                          |         |
| Chlorobium.   | OC        | Chlorobium.  |           |                          |         |
| NCBITaxID=1097;   | OX        | NCBI-TaxID=1097;   |           |                          |         |
| [1]   | RN        | SEQUENCE FROM N.A.   |           |                          |         |
| STRAIN=TLS / ATCC 43652 / DSM 12025;  | RP        |  |           |                          |         |
| MEDLINE=221036901; PubMed=1203901;  | RX        |  |           |                          |         |
| Eisen, J.A.; Nelson, K.E.; Paulsen, I.T.; Heidelberg, J.F.; Wu, M.,   | RA        |  |           |                          |         |
| Dodson, R.J.; Debey, R.; Gruhn, M.L.; Nelson, W.C.; Haft, D.H.,   | RA        |  |           |                          |         |
| Hickey, E.K.; Peterson, J.D.; Durkin, A.S.; Kolonay, J.L.; Yang, F.,  | RA        |  |           |                          |         |
| Holt, I.; Umayan, L.A.; Mason, T.; Bremer, M.; Shea, T.P.; Parksey, D.,   | RA        |  |           |                          |         |
| Nierman, W.C.; Fieldblyum, T.V.; Hansen, C.L.; Graven, M.B.; Radune, D.,  | RA        |  |           |                          |         |
| Vamathevan, J.; Khouri, H.; White, O.; Gruber, T.M.; Ketchum, K.A.,   | RA        |  |           |                          |         |
| Venter, J.C.; Tettelin, H.; Bryant, D.A.; Fraser, C.M.;   | RA        |  |           |                          |         |
| The complete genome sequence of Chlorobium tepidum TLS, a photosynthetic, anaerobic, green-sulfur bacterium.;   | RT        |  |           |                          |         |
| Proc. Natl. Acad. Sci. U.S.A. 99:9509-9514(2002).   | RL        |  |           |                          |         |
| --!- FUNCTION: Involved in the tonB-independent uptake of proteins (By similarity).   | CC        |  |           |                          |         |
| --!- SUBCELLULAR LOCATION: Periplasmic (potential).   | CC        |  |           |                          |         |
| --!- SIMILARITY: Belongs to the tonB family.  | CC        |  |           |                          |         |
| This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to license@isb-sib.ch). | CC        |  |           |                          |         |
| [2]   | RN        | SEQUENCE FROM N.A.   |           |                          |         |
| TISSUE=Extraembryonic tissue;   | RP        |  |           |                          |         |
| McDowell, K.J.; Adams, M.H.; Baker, C.B.;   | RA        |  |           |                          |         |
| Submitted (MAR-1995) to the EMBL/GenBank/DDBJ databases.  | RL        |  |           |                          |         |
| --!- FUNCTION: Transferring are iron binding transport proteins which can bind two atoms of ferric iron in association with the binding of an anion, usually bicarbonate. It is responsible for the transport of iron from sites of absorption and heme degradation to those of storage and utilization. Serum transferrin may also have a further role in stimulating cell proliferation.  | CC        |  |           |                          |         |
| --!- SUBUNIT: Monomer.  | CC        |  |           |                          |         |
| --!- SUBCELLULAR LOCATION: Secreted.  | CC        |  |           |                          |         |
| --!- TISSUE SPECIFICITY: Expressed by the liver and secreted in plasma.   | CC        |  |           |                          |         |
| --!- DOMAIN: Composed of two homologous domains.  | CC        |  |           |                          |         |
| --!- SIMILARITY: Belongs to the transferrin family.   | CC        |  |           |                          |         |
| This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to license@isb-sib.ch). | CC        |  |           |                          |         |
| [3]   | RN        | SEQUENCE FROM N.A.   |           |                          |         |
| TISSUE=Blood;   | RP        |  |           |                          |         |
| DRAMPE; MF_00671; -; 1.   | RA        |  |           |                          |         |
| KW Transport; Protein transport; Periplasmic; Signal; Complete proteome.  | CC        |  |           |                          |         |
| CC SIGNAL   | FT        |  |           |                          |         |
| CC 1 27 Potential.  | PT        |  |           |                          |         |
| CC CHAIN 28 434 ToLB protein.   | PT        |  |           |                          |         |
| CC SEQUENCE 434 AA; 47275 MW; PLA347B89C7A0FF99 CRC64;  | SQ        |  |           |                          |         |
| Query Match 5.0%; Score 93; DB 1; Length 434;   |           |  |           |                          |         |
| Best Local Similarity 21.2%; Pred. No. 2.3; Mismatches 69; Conservative 47; MisMatches 139; Indels 70; Gaps 14;   |           |  |           |                          |         |
| Matches 42  | Qy        |  |           |                          |         |
| 8 FACIMLFGLFLPFPTLRA-B-VEGYIARK-BGASR-TAVLUDKTSANDGKQ----- 58   | Db        |  |           |                          |         |
| 102 FGPLMVLQAGCTKPVYFVRAGLIRACMLVRKAAGGHVYQMAFMKLAALTGTYVYDHLP 101  | Qy        |  |           |                          |         |
| 59 -REWARSLDVTINKGIDFGLFLNLP-A-PLINTINGQNGLNFFASISAYSGDIYAGS 115  | Db        |  |           |                          |         |
| 150 ARRGREIL-----LGPADNFEGCGWRILLAPITAYSOORTGRLGCGCTS 192   | Qy        |  |           |                          |         |
| 116 KRSRGPVLEKHYDSSGKSLLARTYTGERSQLRAIGRLCADLVELLTKRSVFGTRIVF 175   | Db        |  |           |                          |         |
| 193 LTGDRXKN-----VEGEVQVNSTQSTLATEV---NGCVWTFHAGSKTLAGPK 241  | Qy        |  |           |                          |         |
| 176 VANITGKNEIYIMCPDFGENVQLTNSRSISLTPAVSPDGTIYAWTDY--TSGPNLYIKN 233   | Db        |  |           |                          |         |
| 242 PITOMTYNTQDLY---GWOAPPGRASWMPCTC--GSSDLVYVTRADPVIYPRRGPSR 295   | Qy        |  |           |                          |         |
| 234 IATGAKVSVNKHGVCISPAWR-PGTTNTLVLTISYEGDQDLYLI--RADGTVERRITKG 289   | Db        |  |           |                          |         |
| 295 GSLSIISP-----RUVSYLGSSGP 314  | Qy        |  |           |                          |         |





GenCore version 5.1.6  
 Copyright (c) 1993 - 2004 Compugen Ltd.

Om protein - protein search, using sw model

Run on: May 6, 2004, 09:21:36 ; Search time 32.9219 Seconds  
 (without alignments)  
 3373.509 Million cell updates/sec

Title: US-10-650-585-13

Perfect score: 1842  
 Sequence: 1 AHLQWNPPLNVRGGRDAII.....RGWAKAVDPFIPVSEMETTMR 352

Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
 Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
 Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL 25:  
 1: sp\_archea:\*
 2: sp\_bacteria:\*
 3: sp\_fungi:\*
 4: sp\_human:\*
 5: sp\_invertebrate:\*
 6: sp\_mammal:\*
 7: sp\_mhc:\*
 8: sp\_organelle:\*
 9: sp\_phage:\*
 10: sp\_plant:\*
 11: sp\_rabbit:\*
 12: sp\_virus:\*
 13: sp\_vertebrate:\*
 14: sp\_unclassified:\*
 15: sp\_bacteriap:\*
 16: sp\_archeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

### SUMMARIES

| Result No. | Score | Query Match Length | DB ID | Description |
|------------|-------|--------------------|-------|-------------|
| 1          | 1792  | 97.3               | 3010  | 12 Q9DTB6   |
| 2          | 1784  | 96.9               | 3010  | 12 Q9J3F9   |
| 3          | 1783  | 96.8               | 3010  | 12 P90591   |
| 4          | 1782  | 96.7               | 3010  | 12 Q913H7   |
| 5          | 1781  | 96.7               | 3010  | 12 Q9J3H5   |
| 6          | 1780  | 96.6               | 3010  | 12 Q9DTD6   |
| 7          | 1778  | 96.5               | 3008  | 12 Q9J3F4   |
| 8          | 1778  | 96.5               | 3010  | 12 P88801   |
| 9          | 1775  | 96.4               | 3010  | 12 Q68788   |
| 10         | 1775  | 96.4               | 3010  | 12 Q807P3   |
| 11         | 1774  | 96.3               | 3010  | 12 Q81755   |
| 12         | 1774  | 96.3               | 2284  | 12 Q81817   |
| 13         | 1774  | 96.3               | 3010  | 12 P89966   |
| 14         | 1774  | 95.3               | 3010  | 12 Q9J3G6   |
| 15         | 1773  | 96.3               | 3010  | 12 Q99AU2   |
| 16         | 1773  | 96.3               | 3010  | 12 Q9J3H3   |

| RESULT 1 |   |              |      |          |            |
|----------|---|--------------|------|----------|------------|
| Q9DTB6   | Q9DTB6  | PRELIMINARY; | PRT; | 3010 AA. | ALIGNMENTS |
| AC       | Q9DTB6,   |              |      |          |            |
| DT       | 01-MAR-2001 (TREMBLrel. 16, last sequence update)   |              |      |          |            |
| DT       | 01-OCT-2003 (TREMBLrel. 25, last annotation update)   |              |      |          |            |
| DE       | Genome polyprotein.   |              |      |          |            |
| OS       | Hepatitis C virus   |              |      |          |            |
| OS       | Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus   |              |      |          |            |
| OC       | OC  |              |      |          |            |
| OX       | OX  |              |      |          |            |
| RN       | [1]   |              |      |          |            |
| RP       | SEQUENCE FROM N.A.  |              |      |          |            |
| RC       | STRAIN=HCV142; Takanishi K., Iwata R., Matsuoto M., Matsuoto H., Nakao K., Hatahara T., Ohta Y., Kanai K., Maruo H., Baba K., Hijikata M., Misiro S.; Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients with hepatocellular carcinoma: the 1b progression score, revisited; Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases. |              |      |          |            |
| CC       | "Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients with hepatocellular carcinoma: the 1b progression score, revisited;" Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.   |              |      |          |            |
| RT       | - SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MNA (BY SIMILARITY).   |              |      |          |            |
| RI       | "Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients with hepatocellular carcinoma: the 1b progression score, revisited;" Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.   |              |      |          |            |
| CC       | "Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients with hepatocellular carcinoma: the 1b progression score, revisited;" Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.   |              |      |          |            |
| DR       | PIR: A61196; A61196.  |              |      |          |            |
| DR       | PIR: PS0325; PS0329.  |              |      |          |            |
| DR       | HSSP; P26653; IUPP.   |              |      |          |            |
| DR       | GO; GO:0015021; C:integral membrane; IEA.   |              |      |          |            |
| CC       | PROTEIN C AND MNA (BY SIMILARITY).  |              |      |          |            |
| DR       | EMBL: BA04091; BAB18804.1; -.   |              |      |          |            |
| DR       | PIR: A61196.  |              |      |          |            |
| DR       | PIR: PS0325; PS0329.  |              |      |          |            |
| DR       | HSSP; P26653; IUPP.   |              |      |          |            |
| DR       | GO; GO:0015021; C:integral membrane; IEA.   |              |      |          |            |
| DR       | GO; GO:0019028; C:viral capsid; IEA.  |              |      |          |            |
| DR       | GO; GO:0019301; C:viral envelope; IEA.  |              |      |          |            |
| DR       | GO; GO:000524; F:ATP binding; IEA.  |              |      |          |            |
| DR       | GO; GO:000826; F:ATP dependent helicase activity; IEA.  |              |      |          |            |
| DR       | GO; GO:0005489; F:electron transporter activity; IEA.   |              |      |          |            |
| DR       | GO; GO:0005787; F:hydroxole activity; IEA.  |              |      |          |            |
| DR       | GO; GO:001723; F:RNA binding; IEA.  |              |      |          |            |
| DR       | GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  |              |      |          |            |
| GO       | GO:0008236; F:serine-type peptidase activity; IEA.  |              |      |          |            |

MEL MAY / 13:37:04 2004

11-07-03-883-883-13.TBPG

|         |  |              |              |  |                      |
|---------|--|--------------|--------------|--|----------------------|
| DR      | GO; GO:0005198; F:structural molecule activity; IEA.                 | Q9J3F9       | PRELIMINARY; | PRM;   | 3010 AA.             |
| DR      | GO; GO:0006118; P:electron transport; IEA.                           | Q9J3F9       |              |  |                      |
| DR      | GO; GO:000508; P:proteolysis and peptidolysis; IEA.                  | Q9J3F9       |              |  |                      |
| DR      | GO; GO:008350; P:transcription; IEA.                                 | Q9J3F9       |              |  |                      |
| DR      | GO; GO:0019079; P:viral genome replication; IEA.                     | Q9J3F9       |              |  |                      |
| DR      | GO; GO:019087; P:viral transformation; IEA.                          | Q9J3F9       |              |  |                      |
| DR      | InterPro; IPR000003; Cys_Ser_trypsin.                                | Q9J3F9       |              |  |                      |
| DR      | InterPro; IPR00345; Cys_Ser_trypsin.                                 | Q9J3F9       |              |  |                      |
| DR      | InterPro; IPR0110; DEAD.   | Q9J3F9       |              |  |                      |
| DR      | InterPro; IPR02522; HCV_capsid.                                      | Q9J3F9       |              |  |                      |
| DR      | InterPro; IPR0251; HCV_core.   | Q9J3F9       |              |  |                      |
| DR      | InterPro; IPR02511; HCV_env.   | Q9J3F9       |              |  |                      |
| DR      | InterPro; IPR00531; HCV_NS1.   | Q9J3F9       |              |  |                      |
| DR      | InterPro; IPR00518; HCV_NS2.   | Q9J3F9       |              |  |                      |
| DR      | InterPro; IPR00745; HCV_NS4.   | Q9J3F9       |              |  |                      |
| DR      | InterPro; IPR00747; HCV_NS4B.  | Q9J3F9       |              |  |                      |
| DR      | InterPro; IPR02868; HCV_NS5A.  | Q9J3F9       |              |  |                      |
| DR      | InterPro; IPR02166; HCV_RdRp.  | Q9J3F9       |              |  |                      |
| DR      | InterPro; IPR01650; Helicase_C.                                      | Q9J3F9       |              |  |                      |
| DR      | InterPro; IPR00109; Peptidase_C29.                                   | Q9J3F9       |              |  |                      |
| DR      | InterPro; IPR007095; RNA_pol_Ds_ps.                                  | Q9J3F9       |              |  |                      |
| DR      | InterPro; IPR007094; RNA_Pol_PsVir.                                  | Q9J3F9       |              |  |                      |
| DR      | InterPro; IPR01490; HCV_Ns4.   | Q9J3F9       |              |  |                      |
| DR      | PFam; PF01542; HCV_core; 1.  | Q9J3F9       |              |  |                      |
| DR      | PFam; PF01539; HCV_env; 1.   | Q9J3F9       |              |  |                      |
| DR      | PFam; PF01560; HCV_NS1; 1.   | Q9J3F9       |              |  |                      |
| DR      | PFam; PF01538; HCV_NS2; 1.   | Q9J3F9       |              |  |                      |
| DR      | PFam; PF02007; HCV_NS3; 1.   | Q9J3F9       |              |  |                      |
| DR      | PFam; PF01006; HCV_NS4; 1.   | Q9J3F9       |              |  |                      |
| DR      | PFam; PF01001; HCV_NS4b; 1.  | Q9J3F9       |              |  |                      |
| DR      | PFam; PF01506; HCV_NS5a; 1.  | Q9J3F9       |              |  |                      |
| DR      | PFam; PP00271; helicase_C; 1.  | Q9J3F9       |              |  |                      |
| DR      | PFam; PF01598; Viral_RdRp; 1.  | Q9J3F9       |              |  |                      |
| DR      | PFam; PD18602; HCV_NS1; 1.   | Q9J3F9       |              |  |                      |
| DR      | SMART; SM00490; HELICO; 1.   | Q9J3F9       |              |  |                      |
| DR      | PROSITE; PS00090; CYTOCHROME_C; 1.                                   | Q9J3F9       |              |  |                      |
| DR      | ATP-binding; Coat_protein; Envelope_protein; Glycoprotein; Helicase; | Q9J3F9       |              |  |                      |
| DR      | Hydrolase; Nonstructural_protein; Polypeptide;                       | Q9J3F9       |              |  |                      |
| DR      | RNA-directed_RNA_Polymerase; Transferase; Transmembrane;             | Q9J3F9       |              |  |                      |
| DR      | SEQUENCE_ENCODING_3010_AA; 327042_MW; 3807DC6879684C95_CRC64;        | Q9J3F9       |              |  |                      |
| Query   | Match  | 97.3%        | Score        | 1792   | DB 12; Length 3010;  |
| Best    | Local Similarity   | 95.7%        | Pred.        | No. 6_3e-143;  |                      |
| Matches | 337;   | Conservative | 9;           | Mismatches   | 6; Indels 0; Gaps 0; |
| QY      | 1 AHLQVWIPPLNVRGGDAILITCAVHPPELFDTIKULLAIFGPLMLQAGIKTVVFYFVR         | 60           | DR           | GO; GO:0001968; F:RNA-directed RNA_Polymerase_activity; IEA. |                      |
| QY      | 855 AHLQVWIPPLNVRGGDAILITCAVHPPELFDTIKULLAIFGPLMLQAGIKTVVFYFVR       | 914          | DR           | GO; GO:000236; F:serin-type_Peptidase_activity; IEA.         |                      |
| QY      | 61 AOGGLTRACMLVRAKGAGHRYVQAFMKLAALGTGTYVYVHLPQLQDWMAHGRDILAVAEPV     | 120          | DR           | GO; GO:000598; F:structural_molecule_activity; IEA.          |                      |
| Db      | 915 AOGGLTRACMLVRAKGAGHRYVQAFMKLAALGTGTYVYVHLPQLQDWMAHGRDILAVAEPV    | 974          | DR           | GO; GO:000189; P:electron_transport; IEA.                    |                      |
| Db      | 121 IFSDMVVKIITWGADAAKGDIISGLGVSSARGRELLGPDNFEGQWRGLAPTAESQ          | 180          | DR           | GO; GO:000350; P:proteolysis_and_peptidolysis; IEA.          |                      |
| Db      | 975 VFSDMTEKIIITWGADAAKGDIISGLGVSSARGRELLGPDNFEGQWRGLAPTAESQ         | 1034         | DR           | GO; GO:001079; P:viral_genome_replication; IEA.              |                      |
| QY      | 181 OTRGJLGCITSLSITGRDKNQVGEVQVSTATQSFLATCNCVQCVWPHGSKTKTJAGPK       | 240          | DR           | InterPro; IPR000343; Cys_Ser_trypsin.                        |                      |
| QY      | 1035 QTRGLGCITSLSITGRDKNQVGEVQVSTATQSFLATCNCVQCVWPHGSKTKTJAGPK       | 1094         | DR           | InterPro; IPR01410; DEAD.                                    |                      |
| QY      | 241 GPITQMTNTVDDQDVQAPPAGARSMPTCTCGSSDJYLVTRHADVPPVRRGDSRSILS        | 300          | DR           | InterPro; IPR00745; HCV_NS4.                                 |                      |
| Db      | 1095 GPITQMTNTVDDQDVQAPPAGARSMPTCTCGSSDJYLVTRHADVPPVRRGDSRSILS       | 1154         | DR           | InterPro; IPR00745; HCV_capsid.                              |                      |
| QY      | 301 PRPVSVLKGSQQGPICPSGHGVGIRRAAVCTRGVKAADVTPVSENETMR                | 352          | DR           | InterPro; IPR00745; HCV_core; 1.                             |                      |
| QY      | 1155 PRPVSVLKGSQQGPICPSGHGVGIRRAAVCTRGVKAADVTPVSENETMR               | 1206         | DR           | PFam; PF01539; HCV_env; 1.                                   |                      |
| DR      | PFam; PF01560; HCV_NS1; 1.   |              | DR           | PFam; PF01538; HCV_NS2; 1.                                   |                      |
| DR      | PFam; PF02007; HCV_ns3; 1.   |              | DR           | PFam; PF01506; HCV_NS4; 1.                                   |                      |
| DR      | PFam; PF01542; HCV_core; 1.  |              | DR           | PFam; PF01539; HCV_env; 1.                                   |                      |
| DR      | PFam; PF01560; HCV_NS1; 1.   |              | DR           | PFam; PF00521; helicase_C; 1.                                |                      |

DR SMART; SM00487; DEXDC; 1.  
 DR PROSITE; PS00190; CYTOCHROME C; 1.  
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
 KW Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.  
 SEQUENCE 3010 AA; 327102 MW; 7162C9DB93E6B0C7 CRC64;

Query Match 96.9%; Score 1784; DB 12; Length 3010;  
 Best Local Similarity 95.2%; Pred. No. 3e-142;  
 Matches 335; Conservative 11; Mismatches 6; Indels 0; Gaps 0;  
 RT 1 AHLQWIPPLNVRGRGDAAILTCAVHPLFIDITKILAIIGPMLVQAGITKVYFVR 60  
 855 AHLQWIPPLNVRGRGDAAILTCAVHPLFIDITKILAIIGPMLVQAGITKVYFVR 914  
 QY 61 AQLIRACMLVKRAGGHVQMAFMKLAALTGTYYDHLPDWAHAGRLAVAPEV 120  
 915 AQLIRACMLVKRAGGHVQMAFMKLGALTTGTYYDHLPDWAHAGRLAVAPEV 974  
 QY 121 IFPSMEVKITWGADTAAGCDIISGLGPYSARREELIGPANFEQGOWRLAPITAYQ 180  
 121 IFPSMEVKITWGADTAAGCDIISGLGPYSARREELIGPANFEQGOWRLAPITAYQ 180  
 Db 975 VFSDEMKITIWGADTAACGDIISGLGPYSARREELIGPANFEQGOWRLAPITAYQ 1034  
 DR GO; GO:0019031; C:viral envelope; IEA.  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.  
 DR GO; GO:003723; F:RNA binding; IEA.  
 DR GO; GO:003968; F:RNA-directed RNA polymerase activity; IEA.  
 DR GO; GO:008236; F:serine-type peptidase activity; IEA.  
 DR GO; GO:005198; F:structural molecule activity; IEA.  
 DR GO; GO:0016740; F:transferase activity; IEA.  
 DR GO; GO:006508; P:proteolysis and peptidolysis; IEA.  
 DR GO; GO:006350; P:transcription; IEA.  
 DR GO; GO:0019079; P:viral genome replication; IEA.  
 DR InterPro; IPR00903; Cys\_Ser\_trypsin.  
 DR InterPro; IPR01410; DEAd.  
 DR InterPro; IPR02522; HCV capid.  
 DR InterPro; IPR02521; HCV core.  
 DR InterPro; IPR02531; HCV env.  
 DR InterPro; IPR0025319; HCV NS1.  
 DR InterPro; IPR002518; HCV NS2.  
 DR InterPro; IPR000745; HCV NS4a.  
 DR InterPro; IPR001490; HCV NS4b.  
 DR InterPro; IPR02868; HCV NS5a.  
 DR InterPro; IPR02166; HCV RDRP.  
 DR InterPro; IPR001650; Helicase\_C.  
 DR InterPro; IPR004109; Peptidase\_C29.  
 DR InterPro; IPR007095; RNA\_polymerase.  
 DR InterPro; IPR007094; RNA\_Pol\_PsVir.  
 DR Pfam; PF01542; HCV capsid; 1.  
 DR Pfam; PF01539; HCV NS1; 1.  
 DR Pfam; PF01538; HCV NS2; 1.  
 DR Pfam; PF02907; HCV NS3; 1.  
 DR Pfam; PF01006; HCV NS4a; 1.  
 DR Pfam; PF01506; HCV NS4b; 1.  
 DR Pfam; PF00211; helicase\_C; 1.  
 DR ProDom; PD186062; HCV\_NS1; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
 KW Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.  
 FT CHAIN 1 191 CORE PROTEIN.  
 FT CHAIN 192 383 E1.  
 FT CHAIN 384 809 E2.  
 FT CHAIN 810 1026 NS2.  
 FT CHAIN 1027 1657 NS3.  
 FT CHAIN 1658 1711 NS4A.  
 FT CHAIN 1712 1972 NS4B.  
 FT CHAIN 1973 2419 NS5A.  
 FT CHAIN 2420 3010 NS5B.  
 SEQUENCE 3010 AA; 327102 MW; 7162C9DB93E6B0C7 CRC64;

Query Match 96.9%; Score 1783; DB 12; Length 3010;  
 Best Local Similarity 95.5%; Pred. No. 3.e-142;  
 Matches 336; Conservative 11; Mismatches 5; Indels 0; Gaps 0;  
 RT 1 AHLQWIPPLNVRGRGDAAILTCAVHPLFIDITKILAIIGPMLVQAGITKVYFVR 60  
 855 AHLQWIPPLNVRGRGDAAILTCAVHPLFIDITKILAIIGPMLVQAGITKVYFVR 914  
 QY 61 AQLIRACMLVKRAGGHVQMAFMKLAALTGTYYDHLPDWAHAGRLAVAPEV 120  
 915 AQLIRACMLVKRAGGHVQMAFMKLGALTTGTYYDHLPDWAHAGRLAVAPEV 974  
 DR GO; GO:0019031; C:viral envelope; IEA.  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.  
 DR GO; GO:003723; F:RNA binding; IEA.  
 DR GO; GO:003968; F:RNA-directed RNA polymerase activity; IEA.  
 DR GO; GO:008236; F:serine-type peptidase activity; IEA.  
 DR GO; GO:005198; F:structural molecule activity; IEA.  
 DR GO; GO:0016740; F:transferase activity; IEA.  
 DR GO; GO:006508; P:proteolysis and peptidolysis; IEA.  
 DR GO; GO:006350; P:transcription; IEA.  
 DR GO; GO:0019079; P:viral genome replication; IEA.  
 DR InterPro; IPR00903; Cys\_Ser\_trypsin.  
 DR InterPro; IPR01410; DEAd.  
 DR InterPro; IPR002522; HCV capid.  
 DR InterPro; IPR025319; HCV env.  
 DR InterPro; IPR002531; HCV NS1.  
 DR InterPro; IPR002518; HCV NS2.  
 DR InterPro; IPR000745; HCV NS4a.  
 DR InterPro; IPR001490; HCV NS4b.  
 DR InterPro; IPR02868; HCV NS5a.  
 DR InterPro; IPR02166; HCV RDRP.  
 DR InterPro; IPR001650; Helicase\_C.  
 DR InterPro; IPR004109; Peptidase\_C29.  
 DR InterPro; IPR007095; RNA\_polymerase.  
 DR InterPro; IPR007094; RNA\_Pol\_PsVir.  
 DR Pfam; PF01542; HCV capsid; 1.  
 DR Pfam; PF01539; HCV NS1; 1.  
 DR Pfam; PF01538; HCV NS2; 1.  
 DR Pfam; PF02907; HCV NS3; 1.  
 DR Pfam; PF01006; HCV NS4a; 1.  
 DR Pfam; PF01506; HCV NS4b; 1.  
 DR Pfam; PF00211; helicase\_C; 1.  
 DR ProDom; PD186062; HCV\_NS1; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
 KW Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.  
 FT CHAIN 1 191 CORE PROTEIN.  
 FT CHAIN 192 383 E1.  
 FT CHAIN 384 809 E2.  
 FT CHAIN 810 1026 NS2.  
 FT CHAIN 1027 1657 NS3.  
 FT CHAIN 1658 1711 NS4A.  
 FT CHAIN 1712 1972 NS4B.  
 FT CHAIN 1973 2419 NS5A.  
 FT CHAIN 2420 3010 NS5B.  
 SEQUENCE 3010 AA; 327102 MW; 7162C9DB93E6B0C7 CRC64;

Query Match 96.9%; Score 1783; DB 12; Length 3010;  
 Best Local Similarity 95.5%; Pred. No. 3.e-142;  
 Matches 336; Conservative 11; Mismatches 5; Indels 0; Gaps 0;  
 RT 1 AHLQWIPPLNVRGRGDAAILTCAVHPLFIDITKILAIIGPMLVQAGITKVYFVR 60  
 855 AHLQWIPPLNVRGRGDAAILTCAVHPLFIDITKILAIIGPMLVQAGITKVYFVR 914  
 QY 61 AQLIRACMLVKRAGGHVQMAFMKLAALTGTYYDHLPDWAHAGRLAVAPEV 120  
 915 AQLIRACMLVKRAGGHVQMAFMKLGALTTGTYYDHLPDWAHAGRLAVAPEV 974  
 DR GO; GO:0019031; C:viral envelope; IEA.  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.  
 DR GO; GO:003723; F:RNA binding; IEA.  
 DR GO; GO:003968; F:RNA-directed RNA polymerase activity; IEA.  
 DR GO; GO:008236; F:serine-type peptidase activity; IEA.  
 DR GO; GO:005198; F:structural molecule activity; IEA.  
 DR GO; GO:0016740; F:transferase activity; IEA.  
 DR GO; GO:006508; P:proteolysis and peptidolysis; IEA.  
 DR GO; GO:006350; P:transcription; IEA.  
 DR GO; GO:0019079; P:viral genome replication; IEA.  
 DR InterPro; IPR00903; Cys\_Ser\_trypsin.  
 DR InterPro; IPR01410; DEAd.  
 DR InterPro; IPR002522; HCV capid.  
 DR InterPro; IPR025319; HCV env.  
 DR InterPro; IPR002531; HCV NS1.  
 DR InterPro; IPR002518; HCV NS2.  
 DR InterPro; IPR000745; HCV NS4a.  
 DR InterPro; IPR001490; HCV NS4b.  
 DR InterPro; IPR02868; HCV NS5a.  
 DR InterPro; IPR02166; HCV RDRP.  
 DR InterPro; IPR001650; Helicase\_C.  
 DR InterPro; IPR004109; Peptidase\_C29.  
 DR InterPro; IPR007095; RNA\_polymerase.  
 DR InterPro; IPR007094; RNA\_Pol\_PsVir.  
 DR Pfam; PF01542; HCV capsid; 1.  
 DR Pfam; PF01539; HCV NS1; 1.  
 DR Pfam; PF01538; HCV NS2; 1.  
 DR Pfam; PF02907; HCV NS3; 1.  
 DR Pfam; PF01006; HCV NS4a; 1.  
 DR Pfam; PF01506; HCV NS4b; 1.  
 DR Pfam; PF00211; helicase\_C; 1.  
 DR ProDom; PD186062; HCV\_NS1; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
 KW Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.  
 FT CHAIN 1 191 CORE PROTEIN.  
 FT CHAIN 192 383 E1.  
 FT CHAIN 384 809 E2.  
 FT CHAIN 810 1026 NS2.  
 FT CHAIN 1027 1657 NS3.  
 FT CHAIN 1658 1711 NS4A.  
 FT CHAIN 1712 1972 NS4B.  
 FT CHAIN 1973 2419 NS5A.  
 FT CHAIN 2420 3010 NS5B.  
 SEQUENCE 3010 AA; 327102 MW; 7162C9DB93E6B0C7 CRC64;

Query Match 96.9%; Score 1783; DB 12; Length 3010;  
 Best Local Similarity 95.5%; Pred. No. 3.e-142;  
 Matches 336; Conservative 11; Mismatches 5; Indels 0; Gaps 0;  
 RT 1 AHLQWIPPLNVRGRGDAAILTCAVHPLFIDITKILAIIGPMLVQAGITKVYFVR 60  
 855 AHLQWIPPLNVRGRGDAAILTCAVHPLFIDITKILAIIGPMLVQAGITKVYFVR 914  
 QY 61 AQLIRACMLVKRAGGHVQMAFMKLAALTGTYYDHLPDWAHAGRLAVAPEV 120  
 915 AQLIRACMLVKRAGGHVQMAFMKLGALTTGTYYDHLPDWAHAGRLAVAPEV 974  
 DR GO; GO:0019031; C:viral envelope; IEA.  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.  
 DR GO; GO:003723; F:RNA binding; IEA.  
 DR GO; GO:003968; F:RNA-directed RNA polymerase activity; IEA.  
 DR GO; GO:008236; F:serine-type peptidase activity; IEA.  
 DR GO; GO:005198; F:structural molecule activity; IEA.  
 DR GO; GO:0016740; F:transferase activity; IEA.  
 DR GO; GO:006508; P:proteolysis and peptidolysis; IEA.  
 DR GO; GO:006350; P:transcription; IEA.  
 DR GO; GO:0019079; P:viral genome replication; IEA.  
 DR InterPro; IPR00903; Cys\_Ser\_trypsin.  
 DR InterPro; IPR01410; DEAd.  
 DR InterPro; IPR002522; HCV capid.  
 DR InterPro; IPR025319; HCV env.  
 DR InterPro; IPR002531; HCV NS1.  
 DR InterPro; IPR002518; HCV NS2.  
 DR InterPro; IPR000745; HCV NS4a.  
 DR InterPro; IPR001490; HCV NS4b.  
 DR InterPro; IPR02868; HCV NS5a.  
 DR InterPro; IPR02166; HCV RDRP.  
 DR InterPro; IPR001650; Helicase\_C.  
 DR InterPro; IPR004109; Peptidase\_C29.  
 DR InterPro; IPR007095; RNA\_polymerase.  
 DR InterPro; IPR007094; RNA\_Pol\_PsVir.  
 DR Pfam; PF01542; HCV capsid; 1.  
 DR Pfam; PF01539; HCV NS1; 1.  
 DR Pfam; PF01538; HCV NS2; 1.  
 DR Pfam; PF02907; HCV NS3; 1.  
 DR Pfam; PF01006; HCV NS4a; 1.  
 DR Pfam; PF01506; HCV NS4b; 1.  
 DR Pfam; PF00211; helicase\_C; 1.  
 DR ProDom; PD186062; HCV\_NS1; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
 KW Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.  
 FT CHAIN 1 191 CORE PROTEIN.  
 FT CHAIN 192 383 E1.  
 FT CHAIN 384 809 E2.  
 FT CHAIN 810 1026 NS2.  
 FT CHAIN 1027 1657 NS3.  
 FT CHAIN 1658 1711 NS4A.  
 FT CHAIN 1712 1972 NS4B.  
 FT CHAIN 1973 2419 NS5A.  
 FT CHAIN 2420 3010 NS5B.  
 SEQUENCE 3010 AA; 327102 MW; 7162C9DB93E6B0C7 CRC64;

|      |  |  |      |    |  |
|------|--|--|------|----|--|
| Qy   | 241  | GRITOMYNTNDODLVNGWQAPGARSMPCTCGSSDLYLVTVRHADIVPVERRGDSRSLS | 300  | DR | Pfam; PF01538; HCV NS2; 1.   |
| Db   | 1095   | GPTQMYNTNDODLVNGWQAPGARSMPCTCGSSDLYLVTVRHADIVPVERRGDSRSLS  | 1154 | DR | Pfam; PF0106; HCV NS3; 1.  |
| Qy   | 301  | PRPVSYLKGSGGPLCPSGHAVGIFRAVCTRGVAKAVDVFVPMETMR             | 352  | DR | Pfam; PF0101; HCV NS3; 1.  |
| Db   | 1155   | PRPVSYLKGSGGPLCPSGHAVGIFRAVCTRGVAKAVDVFVPMETMR             | 1206 | DR | Pfam; PF0098; Viral RdRp; 1.   |
|      |  |  |      | DR | ProDom; PD186062; HCV NS1; 1.  |
|      |  |  |      | DR | SMART; SM0087; DEXDC; 1.   |
|      |  |  |      | DR | DR PROTEIN; PS0190; CYTOCHROME C; 1.   |
|      |  |  |      | KW | Coat protein; Envelope protein; Glycoprotein; Nonstructural Protein; Polypeptide; RNA-directed RNA Polymerase; Transferase; Transmembrane. |
|      |  |  |      | SQ | SEQUENCE 3010 AA; 327365 MW; D653F73177PA06 CRC64;   |
| Qy   | 99J3H7   | PRELIMINARY; PRT; 3010 AA.                                 |      | Qy | Query Match 96.7%; Score 1782; DB 12; Length 3010; Best local Similarity 94.9%; Seq. No. 4; e=142; Mismatches 11; Indels 0; Gaps 0;        |
| AC   | 99J3H7;  | 01-OCT-2000 (TREMBrel. 15, Last sequence update)           |      | Db | Matches 334; Conservative 11; Mismatches 7; Indels 0; Gaps 0;  |
| DT   | 01-OCT-2000 (TREMBrel. 15, Last sequence update)         |  |      | Qy | 1 AHLOWWIPLNTRGGDRDILLTCAVPELIDITKULLAIFGPMVLTQAGITKVYFVR 974  |
| DR   | 01-OCT-2003 (TREMBrel. 25, Last annotation update)       |  |      | Db | 855 AHLQWVWPPLNTRGGDRDILLTCAVPELIDITKULLAIFGPMVLTQAGITKVYFVR 914   |
| DE   | Genome polyprotein.                                      |  |      | Qy | 61 AQQLTRAMVLVRKAGHRYVQAFMKIAALGTYVYDHLLQDWHAIGRDIAVAPV 120  |
| OS   | Hepatitis C virus.                                       |  |      | Db | 975 VFSDEMETKIIWTQADTAACGDDITLGIVPSARSGRREVGLGADSFEQGWRLIAPITAVSQ 1034   |
| OC   | Hepatitis C virus.                                       |  |      | Qy | 181 QTRALGIGITSTTGDKKNQYEGVQVSTATQSFLATCVCVWTFPHGSKTKLAGPK 240   |
| CC   | Hepatitis C virus.                                       |  |      | Db | 915 QTGLLGCITSITSGDQNQEVQVSTATQSFLATCVCVWTFPHGSKTKLAGPK 974  |
| CC   | Hepatitis C virus.                                       |  |      | Qy | 121 IFSDMEVKITMGADTAACGDDISGLPSARRGRELIGPADNEFGCGWRILAPITAVSQ 180  |
| CC   | Hepatitis C virus.                                       |  |      | Db | 855 AHLQWVWPPLNTRGGDRDILLTCAVPELIDITKULLAIFGPMVLTQAGITKVYFVR 914   |
| EMBL | AF20756; AAF65946.1; -                                   |  |      | Qy | 61 AQQLTRAMVLVRKAGHRYVQAFMKIAALGTYVYDHLLQDWHAIGRDIAVAPV 120  |
| EMBL | A61196; A61196.  |  |      | Db | 1035 QTGLLGCITSITSGDQNQEVQVSTATQSFLATCVCVWTFPHGSKTKLAGPK 1094  |
| PIR  | P00245; P00246.  |  |      | Qy | 241 GPTQMYNTNDODLVNGWQAPGARSMPCTCGSSDLYLVTVRHADIVPVERRGDSRSLS 300  |
| PIR  | P00804; P00804.  |  |      | Db | 1095 GPTQMYNTNDODLVNGWQAPGARSMPCTCGSSDLYLVTVRHADIVPVERRGDSRSLS 1154  |
| PIR  | PS0329; PS0329.  |  |      | Qy | 301 PRPVSYLKGSGGPLCPSGHAVGIFRAVCTRGVAKAVDVFVPMETMR 352   |
| DR   | PRM dependent helicase activity; IEA.                    |  |      | Db | 1155 PRPVSYLKGSGGPLCPSGHAVGIFRAVCTRGVAKAVDVFVPMETMR 1206   |
| DR   | HSSP; P26653; IUPKP.                                     |  |      |    |  |
| DR   | GO; 0016021; C:integral to membrane; IEA.                |  |      |    |  |
| DR   | GO; 0019028; C:viral capsid; IEA.                        |  |      |    |  |
| DR   | GO; 0019031; C:viral envelope; IEA.                      |  |      |    |  |
| DR   | GO; 0005524; F:ATP binding; IEA.                         |  |      |    |  |
| DR   | GO; 0005026; F:ATP dependent helicase activity; IEA.     |  |      |    |  |
| DR   | GO; 0005489; F:electron transporter activity; IEA.       |  |      |    |  |
| DR   | GO; 0003723; F:RNA binding; IEA.                         |  |      |    |  |
| DR   | GO; 003967; F:RNA-directed RNA polymerase activity; IEA. |  |      |    |  |
| DR   | GO; 0002365; F:serine-type peptidase activity; IEA.      |  |      |    |  |
| DR   | GO; 0005198; F:structural molecule activity; IEA.        |  |      |    |  |
| DR   | GO; 0015740; F:transferrase activity; IEA.               |  |      |    |  |
| DR   | GO; 0003118; F:electron transport; IEA.                  |  |      |    |  |
| DR   | GO; 0000508; F:proteoanalysis and Peptidolysis; IEA.     |  |      |    |  |
| DR   | GO; 0006350; P:transcription; IEA.                       |  |      |    |  |
| DR   | GO; 0019079; P:viral genome replication; IEA.            |  |      |    |  |
| DR   | GO; 0019087; P:viral transformation; IEA.                |  |      |    |  |
| DR   | INTERPRO; IPR00903; CYS Ser trypsin.                     |  |      |    |  |
| DR   | INTERPRO; IPR00345; CYTF_heme_BS.                        |  |      |    |  |
| DR   | INTERPRO; IPR01410; DEAD.                                |  |      |    |  |
| DR   | INTERPRO; IPR000745; HCV capsid.                         |  |      |    |  |
| DR   | INTERPRO; IPR02522; HCV core.                            |  |      |    |  |
| DR   | INTERPRO; IPR02519; HCV-env.                             |  |      |    |  |
| DR   | INTERPRO; IPR02531; HCV NS1.                             |  |      |    |  |
| DR   | INTERPRO; IPR02518; HCV NS2.                             |  |      |    |  |
| DR   | INTERPRO; IPR000745; HCV NS4a.                           |  |      |    |  |
| DR   | INTERPRO; IPR01490; HCV NS4b.                            |  |      |    |  |
| DR   | INTERPRO; IPR02868; HCV NS5a.                            |  |      |    |  |
| DR   | INTERPRO; IPR02816; HCV Rdrp.                            |  |      |    |  |
| DR   | INTERPRO; IPR00109; Peptidase C29.                       |  |      |    |  |
| DR   | INTERPRO; IPR007095; RNA_Pol DS PS.                      |  |      |    |  |
| DR   | INTERPRO; IPR007094; RNA_Pol_PSVir.                      |  |      |    |  |
| DR   | PFam; PF01543; HCV capsid; 1.                            |  |      |    |  |
| DR   | PFam; PF01542; HCV core; 1.                              |  |      |    |  |
| DR   | PFam; PF01539; HCV env; 1.                               |  |      |    |  |
| DR   | PFam; PF01560; HCV NS1; 1.                               |  |      |    |  |

|             |   |    |  |
|-------------|---|----|--|
| DR          | GO: GO:000524; F:ATP binding; IEA.  | Qy | 301 PRPVSYLKGSQQPLCPSCHAvgfRRAVCTRGVAKAVDFIPVESMETMR 352   |
| DR          | GO; GO:000806; F:ATP dependent helicase activity; IEA.  | Db | 1155 PRPVSYLKGSQQPLCPSCHAvgfRRAVCTRGVAKAVDFIPVESMETMR 1206 |
| DR          | GO; GO:0016787; F:electron transporter activity; IEA.   |    |  |
| DR          | GO; GO:0003723; F:RNA binding; IEA.   |    |  |
| DR          | GO; GO:0008336; F:serine-type peptidase activity; IEA.  |    |  |
| DR          | GO; GO:0005298; F:structural molecule activity; IEA.  |    |  |
| DR          | GO; GO:0016740; F:transferase activity; IEA.  |    |  |
| DR          | GO; GO:0006118; F:electron transport; IEA.  |    |  |
| DR          | GO; GO:000508; F:protein and peptidolysis; IEA.   |    |  |
| DR          | GO; GO:0006350; F:transcription; IEA.   |    |  |
| DR          | GO; GO:0019079; F:viral genome replication; IEA.  |    |  |
| DR          | GO; GO:0019887; F:viral transformation; IEA.  |    |  |
| DR          | InterPro; IPR00903; Cys_Ser_trypsin.  |    |  |
| DR          | InterPro; IPR00345; Cytc_heme_BS.   |    |  |
| DR          | InterPro; IPR001410; DEAD.  |    |  |
| DR          | InterPro; IPR02222; HCV_capsid.   |    |  |
| DR          | InterPro; IPR002321; HCV_core.  |    |  |
| DR          | InterPro; IPR00231; HCV_env.  |    |  |
| DR          | InterPro; IPR002318; HCV_NS1.   |    |  |
| DR          | InterPro; IPR000745; HCV_NS4a.  |    |  |
| DR          | InterPro; IPR001490; HCV_NS4b.  |    |  |
| DR          | InterPro; IPR002868; HCV_NS5a.  |    |  |
| DR          | InterPro; IPR002166; HCV_RdRp.  |    |  |
| DR          | InterPro; IPR001650; Helicase_C.  |    |  |
| DR          | InterPro; IPR004109; Peptidase_C29.   |    |  |
| DR          | InterPro; IPR007095; RNA_pol_DS_PS.   |    |  |
| DR          | InterPro; IPR007094; RNA_pol_PsVr.  |    |  |
| DR          | Pfam; PF01533; HCV_capsid; 1.   |    |  |
| DR          | Pfam; PF01542; HCV_core; 1.   |    |  |
| DR          | Pfam; PF01539; HCV_env; 1.  |    |  |
| DR          | Pfam; PF01560; HCV_NS1; 1.  |    |  |
| DR          | Pfam; PF01538; HCV_NS2; 1.  |    |  |
| DR          | Pfam; PF020907; HCV_NS3; 1.   |    |  |
| DR          | Pfam; PF01005; HCV_NS4a; 1.   |    |  |
| DR          | Pfam; PF01001; HCV_NS4b; 1.   |    |  |
| DR          | Pfam; PF01506; HCV_NS5a; 1.   |    |  |
| DR          | Pfam; PF00071; helicase_C; 1.   |    |  |
| DR          | Pfam; PF00998; viral_RdRp; 1.   |    |  |
| DR          | Pfam; PF00062; HCV_NS1; 1.  |    |  |
| DR          | SMART; SM00487; DEAD_C; 1.  |    |  |
| DR          | PROSITE; PS0190; CYTOCHROME_C; 1.   |    |  |
| KW          | ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;  |    |  |
| KW          | Hydrolyase; Nonstructural protein; Polyprotein;   |    |  |
| KW          | RNA-directed RNA polymerase; Transferase; Transmembrane;  |    |  |
| SQ          | SEQUENCE 3010 AA; 326801 MW; 9FE33D1B393B7A4B CRC64;  |    |  |
| Query Match | 95.7%; Score 1781; DB 12; length 3010; Best Local Similarity 95.5%; Pred. No. 5.6-147; Matches 336; Conservative 8; Mismatches 8; Indels 0; Gaps 0; |    |  |
| Qy          | 1 AHLQWVPIPLNVRGGDAILITCAVHPLFIDITKLLAICPMLVQAGITKVYFVR 60  |    |  |
| Db          | 855 AHQWVPPVLNVGGRDAILITCAVHPLFIDITKLLAICPMLVQAGITRVYFVR 914  |    |  |
| Qy          | 61 A QGII RACM VVKRAGGHVQMAFMKLAALTGTVYDHLTPLODWAHAGRLAVAEPV 120  |    |  |
| Db          | 915 A QGII RACM VVKRAGGHVQMAFMKLAALTGTVYDHLTPRWAHAGRLAVAEPV 974   |    |  |
| Qy          | 121 IFSDMEVKLITWGDAATTAAAGDTISGLPLYSARGRGELIIGPNDNEFGOGWRLLAPITYSQ 180  |    |  |
| Db          | 975 VFSDEMETKLTIWGDAATTAAAGDTISGLPLYSARGRGELIIGPNDNEFGOGWRLLAPITYSQ 1034  |    |  |
| Qy          | 181 QTRGLGCGTTSLSGRDKNQVEGVGVQVSTASFLATCVNQVWTFHGASSKLAGPK 240  |    |  |
| Db          | 1035 QTRGLGCGTTSLSLGRDKNQVEGVQVQVSTASFLATCVNQVWTFHGASSKLAGPK 1094   |    |  |
| Qy          | 241 GPTOMTYNTDQDLVGWQAPPGARSMTPTCGSSDLYVTRADIVPYRRRSRSGLS 300   |    |  |
| Db          | 1095 GP1QMYTINVDDLDLVGQAPPGARSMTPTCGSSDLYLVTRADIVPYRRRSRSGLS 1154   |    |  |

Query Match 96.6%; Score 1780; DB 12; Length 3010;  
Best Local Similarity 95.7%; Pred. No. 5. 6.e-142; Indels 0; Gaps 0;  
Matches 337; Conservative 7; Mismatches 8; Index 0; Gaps 0;  
Score 1780; DB 12; Length 3010;  
Best Local Similarity 95.7%; Pred. No. 5. 6.e-142; Indels 0; Gaps 0;  
Matches 337; Conservative 7; Mismatches 8; Index 0; Gaps 0;

QY 1 AHQWVWPLNTRGGDAILLTCAVPHELFDITKLLAIFGPMVLQAGITKPYFVR 60  
Db 855 AHQWVWPLNTRGGDAILLTCAVPHELFDITKLLAIFGPMVLQAGITKPYFVR 914  
QY 61 AQGLIRACMLVRKAAGGHYQVQAFMKLAALGTGTYVVDHTLPLQDWAHAGRLDVAVEPV 120  
915 AQGLIRACMLVRKAAGGHYQVQAFMKLAALGTGTYVVDHTLPLQDWAHAGRLDVAVEPV 974  
QY 121 IFSDMVEKITWGADTAACGDTSIGLVSARSGRETLGPADNEFGQWRILAPITAYSQ 180  
975 VFSDEMETKITTWGADTAACGDTSIGLVSARSGRETLGPADNEFGQWRILAPITAYSQ 1034

QY 181 QTRGLLGCCITSITGRDKDNQVEGVQVSTATQSFLATCNGVCMVPHGAGSKTLAGPK 240  
1035 QTRGLLGCCITSITGRDKDNQVEGVQVSTATQSFLATCNGVCMVPHGAGSKTLAGPK 1094

QY 241 GPTQMYINVDQDLVQGAPGARSMPCTGSSDLYLVTRADVIVPVRGGDSRSILS 300  
QY 1095 GRPTQMYINVDQDLVQGAPGARSMPCTGSSDLYLVTRADVIVPVRGGDSRSILS 1154

QY 301 PRVSYLKGASSGAPLCPSGHAGVIFRAVTRGVAKVADIPVSMETMR 352  
QY 1155 PRVSYLKGASSGAPLCPSGHAGVIFRAVTRGVAKVADIPVSMETMR 1206

RESULT 7

Q9J3F4 PRELIMINARY; PRT; 3008 AA.  
ID Q9J3F4; PRELIMINARY; PRT; 3008 AA.  
AC 09J3F4; 2000 (Tremblrel. 15, Last sequence update)  
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
DE Genome polyprotein.

GN MD34.  
OS Hepatitis C virus.  
OC Viruses; SRNA, positive-strand viruses, no DNA stage; Flaviviridae;  
OC Hepadivirus.  
OX NCBI\_TaxId:1103;  
RN SEQUENCE FROM N.A.

RC STRAIN:MD34;  
RA Nagayama K., Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;  
RT "Characteristics of hepatitis C viral genome associated with disease progression," Submitted (Nov-1999) to the EMBL/GenBank/DDBJ databases.  
RL -  
CC -  
CC -  
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA (BY SIMILARITY).  
CC PROTEIN C AND MRNA (BY SIMILARITY).  
DR EMBL; AF205024; AAF61205.1; -  
DR PIR; A61196; A61196.  
DR PQ2046; PQ2046.  
DR PIR; PS0329; PS0329.  
DR HSSP; P26603; IJXPF.

Query Match 95.5%; Score 1778; DB 12; Length 3008;  
Best Local Similarity 95.5%; Pred. No. 9. 8.e-142; Indels 0; Gaps 0;  
Matches 336; Conservative 7; Mismatches 9; Index 0; Gaps 0;

QY 1 AHQWVWPLNTRGGDAILLTCAVPHELFDITKLLAIFGPMVLQAGITKPYFVR 60  
Db 853 AHQWVWPLNTRGGDAILLTCAVPHELFDITKLLAIFGPMVLQAGITKPYFVR 912  
QY 61 AQGLIRACMLVRKAAGGHYQVQAFMKLAALGTGTYVVDHTLPLQDWAHAGRLDVAVEPV 120  
913 AQGLIRACMLVRKAAGGHYQVQAFMKLAALGTGTYVVDHTLPLQDWAHAGRLDVAVEPV 972

QY 121 IFSDMVEKITWGADTAACGDTSIGLVSARSGRETLGPADNEFGQWRILAPITAYSQ 180  
973 VFSDEMETKITTWGADTAACGDTSIGLVSARSGRETLGPADNEFGQWRILAPITAYSQ 1032

QY 181 QTRGLLGCCITSITGRDKDNQVEGVQVSTATQSFLATCNGVCMVPHGAGSKTLAGPK 240  
1033 QTRGLLGCCITSITGRDKDNQVEGVQVSTATQSFLATCNGVCMVPHGAGSKTLAGPK 1094

QY 241 GPTQMYINVDQDLVQGAPGARSMPCTGSSDLYLVTRADVIVPVRGGDSRSILS 300

DR GO; GO:0016021; C:integral to membrane; IEA.  
DR GO; GO:001928; C:viral capsid; IEA.  
DR GO; GO:001931; C:viral envelope; IEA.  
DR GO; GO:000524; F:ATP binding; IEA.  
DR GO; GO:000884; F:electron transporter activity; IEA.  
DR GO; GO:00589; F:electron transport; IEA.  
DR GO; GO:001687; F:hydrolase activity; IEA.  
DR GO; GO:000323; F:RNA binding; IEA.  
DR GO; GO:000368; F:RNA-directed RNA polymerase activity; IEA.  
DR GO; GO:000526; F:serine-type peptidase activity; IEA.  
DR GO; GO:000519; F:structural molecule activity; IEA.  
DR GO; GO:001907; F:transferase activity; IEA.  
DR GO; GO:001908; F:viral genome replication; IEA.  
DR GO; GO:001979; F:viral transformation; IEA.  
DR InterPro; IPR00903; Cys Ser trypsin.  
DR InterPro; IPR00345; Cyt\_C\_heme\_BS.  
DR InterPro; IPR0141; DEAD.  
DR InterPro; IPR002322; HCV capsid.  
DR InterPro; IPR00521; HCV core.  
DR InterPro; IPR002519; HCV env.  
DR InterPro; IPR002331; HCV\_N51.  
DR InterPro; IPR002518; HCV\_N52.  
DR InterPro; IPR001745; HCV\_N54a.  
DR InterPro; IPR001490; HCV\_N54b.  
DR InterPro; IPR002868; HCV\_N55.  
DR InterPro; IPR00166; HCV\_RdRP.  
DR InterPro; IPR01550; Helicase\_C.  
DR InterPro; IPR004109; Peptidase\_C29.  
DR InterPro; IPR007095; RNA\_Pol\_DS\_PS.  
DR InterPro; IPR007094; RNA\_Pol\_PstI.  
DR Pfam; PF01533; HCV\_capsid; 1.  
DR Pfam; PF01539; HCV\_core; 1.  
DR Pfam; PF01560; HCV\_N51; 1.  
DR Pfam; PF01538; HCV\_N52; 1.  
DR Pfam; PF02917; HCV\_N53; 1.  
DR Pfam; PF01001; HCV\_N5b; 1.  
DR Pfam; PF01536; HCV\_N5a; 1.  
DR Pfam; PF02171; helicase\_C; 1.  
DR Pfam; PF00998; viral\_RdRP; 1.  
DR Prodom; PD16062; HCV\_N51; 1.  
DR SMART; SM00887; DEXDC; 1.  
DR PROSITE; PS00190; CYTOCHROME\_C; 1.  
KW ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;  
KW RNA-directed RNA polymerase; Transferase; Transmembrane;  
SEQUENCE 3008 AA; MW; 994E09E14C3109F4 CRC64;





DR Prodrom; PDB186062; HCV\_N51; 1.  
 DR SMART; SMO047; DEXDC; 1.  
 DR PROSITE; PS00190; CYTOCHROME\_C; 1.  
 KW SEQUENCE 3010 AA; 327097 MW; BE6418C7A723E686 CRC64;

Query Match 95.4%; Score 1775; DB 12; Length 3010;  
 Best Local Similarity 95.2%; Pred. No. 1.8e-141; RT  
 Matches 335; Conservative 8; Mismatches 9; Indels 0; Gaps 0; RT  
 Query 1 AHLQWIPPLNVRGERDAITLTCAVHPELFDITKULLAIFGQLMVHQAGITKVPPFVR 60  
 Db 855 AHLQWIPPLNVRGERDAITLTCAVHPELFDITKULLAIFGQLMVHQAGITKVPPFVR 914  
 Query 61 AQLIRACMLVRKAAGGHYVQAFMFLALGTGTVYDHLPDWAHAGRLDVLAVEPV 120  
 Db 915 AQLIRACMLVRKAAGGHYVQAFMFLALGTGTVYDHLPDWAHAGRLDVLAVEPV 974  
 Query 121 IFSMEMVKLTIWGAUTAAAGDDISGLPSARREBILGPAFFEGQWRLLAPITYSQ 180  
 Db 975 VFSDMETKLTIWGAUTAAAGDDISGLPSARREBILGPAFFEGQWRLLAPITYSQ 240  
 Query 181 QTRGLGCTTSLTERDKRQVVEGVQVSTATOSFLATCNGVQWTGASSKTLGPK 240  
 Db 1035 QTRGLGCTTSLTERDKRQVVEGVQVSTATOSFLATCNGVQWTGASSKTLGPK 1094  
 Query 241 GPITOMYTIVNDQDLUGWQAPPGARSMTCTCGSSDLYLVTRADVIYRGRDSRGSLIS 300  
 Db 1095 GPITOMYTIVNDQDLUGWQAPPGARSMTCTCGSSDLYLVTRADVIYRGRDSRGSLIS 1154  
 Query 301 PRPVSYLKGSSSGGPILCPSPGHAVGIFRAAVCTRVAKAVDFIPIVESMMTTMR 352  
 Db 1155 PRPVSYLKGSSSGGPILCPSPGHAVGIFRAAVCTRVAKAVDFIPIVESMMTTMR 1206  
 RESULT 11  
 ID Q81755 PRELIMINARY; PRT: 1186 AA.  
 AC Q81755;  
 DT 01-NOV-1996 (Tremblrel. 01, Created)  
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
 DE Polyprotein (Fragment).  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Herpesviridae;  
 RN [1] NCBI\_TAXID=11103;  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=9108850; PubMed=2175903;  
 RA Kato N., Hijikata M., Ootsuyama Y., Nakagawa M., Ohkoshi S.,  
 RA Sugimura T., Shimotohno K.;  
 RT "Molecular cloning of the human hepatitis C virus genome from Japanese  
 patients with non-A, non-B hepatitis";  
 RL Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528 (1990).  
 RN [2] SEQUENCE FROM N.A.  
 RP MEDLINE=9204440; PubMed=1658196;  
 RA Okamoto H., Okada S., Sugiyama Y., Kurai K., Iizuka H., Machida A.,  
 RA Miyakawa Y., Mayumi M.;  
 RT "Nucleotide sequences of the genomic RNA of hepatitis C virus isolated  
 from a human carrier: comparison with reported isolates for conserved  
 and divergent regions";  
 RT J. Gen. Virol. 72:2697-2704 (1991).  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=1140098; PubMed=1847440;  
 RA Andoh T., Yoshida T., Okayama H.;  
 RT "The structure and organization of the Hepatitis C virus genome  
 isolated from human carriers";  
 RL J. Virol. 65:1105-1113 (1991).

RN [4] SEQUENCE FROM N.A.  
 RP Sequence FROM N.A., PubMed=1314449;  
 RX MEDLINE=9230206; PubMed=1848704;  
 RA Choo Q.-L., Richman K., Han J.-H., Berger K., Lee C., Dong C.,  
 Gallegos C., Coit D., Medina Selby A., Barr P.J., Weiner A.,  
 Bradley D.W., Kuo G., Houghton M.;  
 RT "Genetic organization and diversity of the hepatitis C virus.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455 (1991);  
 RN [5] SEQUENCE FROM N.A., PubMed=1314449;  
 RP MEDLINE=9230206; Chen P., Lin M., Tai K., Liu P., Lin C., Chen D.;  
 RA Chen P., Lin M., Tai K., Liu P., Lin C., Chen D.;  
 RT "Sequence determination and antigenomic RNA";  
 RL "The Taiwanese hepatitis C virus genome: Sequence determination and  
 mapping the 5' termini of viral genomic and antigenomic RNA";  
 RN [6] SEQUENCE FROM N.A., PubMed=1314459;  
 RX MEDLINE=9230222; Okamoto H., Kurai K., Okada S., Yamamoto K., Iizuka H., Tanaka T., Fukuda S., Tsuda F., Misiro S.;  
 RT "Full-length sequence of a hepatitis C virus genome having poor  
 homology to reported isolates: Comparative study of four distinct  
 genotypes";  
 RT "Putative nonstructural precursor protein of hepatitis C virus.";  
 RL Virology 188:331-341 (1992).  
 RN [7] SEQUENCE FROM N.A., PubMed=9323208;  
 RX MEDLINE=9323208; PubMed=8392605;  
 RA Hijikata M., Mizushima H., Akegi T., Mori S., Kakiuchi N., Kato N., Tanaka T., Kimura K., Shimotohno K.;  
 RT "Two distinct protease activities required for the processing of a  
 putative nonstructural precursor protein of hepatitis C virus.";  
 RL Virology 167:4655-4675 (1993).  
 RN [8] SEQUENCE FROM N.A., PubMed=9323208;  
 RA Hijikata M.;  
 RL Submitted (DEC-1993) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; D1197; BAA20075.1; -;  
 DR PIR; A61196; A61196.  
 DR PIR; PS0329; PS0329.  
 DR PDB; 1DXP; 28-MAR-02.  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0005026; F:ATP dependent helicase activity; IEA.  
 DR GO; GO:0016787; F:hydrolyase activity; IEA.  
 DR GO; GO:003676; F:nucleic acid binding; IEA.  
 DR GO; GO:000236; F:serine-type peptidase activity; IEA.  
 DR GO; GO:000508; P:proteinase and peptidolytic; IEA.  
 DR InterPro; IPR00903; Cys\_Ser\_trypsin.  
 DR InterPro; IPR00410; DEAD.  
 DR InterPro; IPR002518; HCV NS2.  
 DR InterPro; IPR00745; HCV NS4a.  
 DR InterPro; IPR001490; HCV NS4b.  
 DR InterPro; IPR001650; Helicase\_C.  
 DR InterPro; IPR004109; Peptidase\_C29.  
 DR Pfam; PF01538; HCV NS2; 1.  
 DR Pfam; PF00907; HCV NS3; 1.  
 DR InterPro; IPR002518; HCV NS4a; 1.  
 DR Pfam; PF0106; HCV NS4b; 1.  
 DR Pfam; PF0101; HCV NS4b; 1.  
 DR Pfam; PF0021; helicase\_C; 1.  
 DR SMART; SMO0487; DEXDC; 1.  
 KW ATP-binding; Helicase; Hydrolase.  
 FT NON\_TER 1  
 SQ SEQUENCE 1186 AA; 12680 MW; 34170478BA23729A CRC64;  
 Query Match 96.3%; Score 1774; DB 12; Length 1186;  
 Best Local Similarity 95.2%; Pred. No. 6.4e-142; RT  
 Matches 335; Conservative 10; Mismatches 7; Indels 0; Gaps 0; RT  
 Query 1 AHLQWIPPLNVRGERDAITLTCAVHPELFDITKULLAIFGQLMVHQAGITKVPPFVR 60  
 Db 133 AHLQWIPPLNVRGERDAITLTCAVHPELFDITKULLAIFGQLMVHQAGITKVPPFVR 192  
 Query 61 AQLIRACMLVRKAAGGHYVQAFMFLALGTGTVYDHLPDWAHAGRLDVLAVEPV 120

|                  |             |   |      |  |
|------------------|-------------|---|------|--|
| Db               | 193         | AQGLIRACMLVRKVAGGHYVQAMFKLALITGTGIVDILPRLDWAHGIRDVLAVEPV  | 252  | DR; GO:0006508; P:proteolysis and peptidolysis; IEA. |
| Qy               | 121         | IFSDMEVKITLTWGAATACCDISQLPNSARRGREIILGPADNFEQGWRLLAPITAYSQ  | 180  | GO; GO:0006350; P:transcription; IEA.                |
| Db               | 253         | VFSDEMETKILTWGAATACCDISQLPNSARRGREIILGPADNFEQGWRLLAPITAYSQ  | 312  | DR; GO:0009079; P:viral genome replication; IEA.     |
| Qy               | 181         | QTREGLGICITSLTERKDKNOVEGEVQVSTATOSFLATCNGVWVHAGSKTLAGPK   | 240  | DR; GO:0019087; P:viral transformation; IEA.         |
| Db               | 313         | QTREGLGICITSLTERKDKNOVEGEVQVSTATOSFLATCNGVWVHAGSKTLAGPK   | 372  | DR; InterPro; IPR00903; Cys Ser_trypsin.             |
| Qy               | 241         | GPIOTMNTDQDUGWQAPPGARSMTPTCGSSDLVLYLUTRAADVIPVRGRGSRSGLS  | 300  | DR; InterPro; IPR001410; DEAD.                       |
| Db               | 373         | GPIOTMNTDQDUGWQAPPGARSMTPTCGSSDLVLYLUTRAADVIPVRGRGSRSGLS  | 432  | DR; InterPro; IPR002518; HCV NS2.                    |
| Qy               | 301         | PRPVSYLKGSGGGPUCPSCHVVGIFRAAVCTRGVAKAVDFIPVESMETTMR   | 352  | DR; InterPro; IPR000745; HCV NS4b.                   |
| Db               | 433         | PRPVSYLKGSGGGPUCPSCHVVGIFRAAVCTRGVAKAVDFIPVESMETTMR   | 484  | DR; InterPro; IPR01490; HCV NS4b.                    |
| <b>RESULT 12</b> |             |   |      |  |
| Q81817           |             | PRELIMINARY;  | PRT; | 2284 AA.   |
| AC               | Q81817;     |   |      |  |
| DT               | 01-NOV-1996 | (TREMBrel. 01, Created)   |      |  |
| DT               | 01-OCT-2003 | (TREMBrel. 25, last annotation update)  |      |  |
| DT               |             | Polypeptide precursor (Genome polypeptide).   |      |  |
| OS               |             | Heptatitis C virus type 2.  |      |  |
| OC               |             | Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;   |      |  |
| OX               |             | Hpacivirus.   |      |  |
| NCBI-TaxID       |             | NCBI-TaxID=40271;   |      |  |
| RN               |             | [1]   |      |  |
| RP               |             | SEQUENCE FROM N.A.  |      |  |
| RX               |             | MEDLINE=9406847; PubMed=7504283;  |      |  |
| RA               |             | Hijikata M., Mizushima H., Tanji Y., Komada Y., Hirowatari Y.,  |      |  |
| RA               |             | Akagi T., Kimura K., Shimotohno K.;   |      |  |
| RT               |             | "Proteolytic processing and membrane association of putative nonstructural proteins of hepatitis C virus.";         |      |  |
| RT               |             | Proc. Natl. Acad. Sci. U.S.A. 90:10773-10777(1993).   |      |  |
| RN               |             | [2]   |      |  |
| RP               |             | SEQUENCE FROM N.A.  |      |  |
| RX               |             | MEDLINE=9433810; PubMed=8056334;  |      |  |
| RA               |             | Tanji Y., Hijikata M., Hirowatari Y., Shimotohno K.;  |      |  |
| RT               |             | "Identification of the domain required for trans-cleavage activity of hepatitis C viral serine proteinase.";        |      |  |
| RT               |             | Gene 145:215-219(1994).   |      |  |
| RN               |             | [3]   |      |  |
| RP               |             | SEQUENCE FROM N.A.  |      |  |
| RX               |             | MEDLINE=95056078; PubMed=7966638;   |      |  |
| RA               |             | Tanji Y., Hijikata M., Hirowatari Y., Shimotohno K.;  |      |  |
| RT               |             | "Hepatitis C virus polyprotein processing: kinetics and mutagenic analysis of serine protease-dependent cleavage."; |      |  |
| RT               |             | J. Virol. 68:8418-8422(1994).   |      |  |
| RL               |             | [4]   |      |  |
| RN               |             | SEQUENCE FROM N.A.  |      |  |
| RP               |             | MEDLINE=95056078; PubMed=7953491;   |      |  |
| RA               |             | Tanji Y., Hijikata M., Satch S., Kaneko T., Shimotohno K.;  |      |  |
| RT               |             | "Hepatitis C virus-encoded nonstructural protein NS4A has versatile functions in viral protein processing.";        |      |  |
| RT               |             | J. Virol. 69:1575-1581(1995).   |      |  |
| RL               |             | EMBL: D16435; BAA33905.1; -.  |      |  |
| DR               |             | PIR: A61196; A61196.  |      |  |
| DR               |             | PIR: P00246; P00246.  |      |  |
| DR               |             | PIR: PS329; PS329.  |      |  |
| DR               |             | HSSP: P26663; 1JXP.   |      |  |
| DR               |             | GO; GO:0019012; C:virion; IEA.  |      |  |
| DR               |             | GO; GO:0005524; F:ATP binding; IEA.   |      |  |
| DR               |             | F:ATP dependent helicase activity; IEA.   |      |  |
| DR               |             | GO; GO:000826; F:RNA binding; IEA.  |      |  |
| DR               |             | GO; GO:000323; F:RNA binding; IEA.  |      |  |
| DR               |             | GO; GO:0003648; F:RNA-directed RNA polymerase activity; IEA.  |      |  |
| DR               |             | GO; GO:0008256; F:serine-type peptidase activity; IEA.  |      |  |
| DR               |             | GO; GO:0016740; P:transferase activity; IEA.  |      |  |
| <b>RESULT 13</b> |             |   |      |  |
| P89956           |             | PRELIMINARY;  | PRT; | 3010 AA.   |
| P89956           |             |   |      |  |
| AC               |             |   |      |  |
| P89956           |             | 01-MAY-1997 (TREMBrel. 03, Created)   |      |  |
| DT               |             | 01-OCT-2003 (TREMBrel. 25, last annotation update)  |      |  |
| DR               |             | RNA for polyprotein (Genome polyprotein).   |      |  |

|                   | Matches | Conservative  | 10;   | Mismatches | 7;   | Indels | 0;  | Gaps   |       |
|-------------------|---------|---|-------|------------|------|--------|-----|--------|-------|
| OC                | 1       | AHLQWQIPLANRGRGDRDAILITLCAVHBLIFPQLMVQAGITKVPFVR  | 60    |            |      |        |     |        |       |
| OC                | 855     | AHQTYWPINLVNGVGRDAILITLCVHBLIFPQLMVQAGITKVPFVR  | 914   |            |      |        |     |        |       |
| RC                | 61      | AQGLTRACMVLRKAGGHVQMAFMKAALTGTGVYDHPLQDWAHGLRDIAVAPV  | 120   |            |      |        |     |        |       |
| NCBI_TaxID=1103;  |         | AQGLTRACMVLRKAGGHVQMAFMKAALTGTGVYDHPLQDWAHGLRDIAVAPV  | 974   |            |      |        |     |        |       |
| RA                | 111     | SEQUENCE FROM N.A.  |       |            |      |        |     |        |       |
| RA                | 1       | SIRAI-type 1b;  |       |            |      |        |     |        |       |
| RL                | 1       | Tanaka T.;  |       |            |      |        |     |        |       |
| RN                | 1       | Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.   |       |            |      |        |     |        |       |
| RP                | 1       | [2]   |       |            |      |        |     |        |       |
| RC                | 1       | SEQUENCE FROM N.A.  |       |            |      |        |     |        |       |
| RA                | 1       | STRAY-type 1b;  |       |            |      |        |     |        |       |
| RT                | 1       | TANAKA T.;  |       |            |      |        |     |        |       |
| RL                | 1       | "TMORF";  |       |            |      |        |     |        |       |
| CC                | 1       | Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases. COVERED BY A  |       |            |      |        |     |        |       |
| CC                | 1       | - SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A  |       |            |      |        |     |        |       |
| CC                | 1       | LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTBINS:  |       |            |      |        |     |        |       |
| CC                | 1       | PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF  |       |            |      |        |     |        |       |
| CC                | 1       | PROTEIN C AND MRNA (BY SIMILARITY).   |       |            |      |        |     |        |       |
| EMBL; D89872;     | 1       | BAA14035.1; -   |       |            |      |        |     |        |       |
| DR                | 1       | PIR; A61196; A61196.  |       |            |      |        |     |        |       |
| DR                | 1       | P00245; P00246.   |       |            |      |        |     |        |       |
| DR                | 1       | PIR; PQ0804; PQ0804.  |       |            |      |        |     |        |       |
| DR                | 1       | PIR; PS0329; PS0329.  |       |            |      |        |     |        |       |
| DR                | 1       | HSSP; P2663; LJXP.  |       |            |      |        |     |        |       |
| GO; GO:016021;    | 1       | C:integral to membrane; IEA.  |       |            |      |        |     |        |       |
| GO; GO:0019028;   | 1       | C:viral capsid; IEA.  |       |            |      |        |     |        |       |
| GO; GO:005522;    | 1       | F:ATP binding; IEA.   |       |            |      |        |     |        |       |
| GO; GO:0005261;   | 1       | F:ATP dependent helicase activity; IEA.   |       |            |      |        |     |        |       |
| GO; GO:000350;    | 1       | F:RNA binding; IEA.   |       |            |      |        |     |        |       |
| GO; GO:003968;    | 1       | F:RNA-directed RNA polymerase activity; IEA.  |       |            |      |        |     |        |       |
| GO; GO:000236;    | 1       | F:serine-type peptidase activity; IEA.  |       |            |      |        |     |        |       |
| GO; GO:0005198;   | 1       | F:structural molecule activity; IEA.  |       |            |      |        |     |        |       |
| GO; GO:0016740;   | 1       | F:transferase activity; IEA.  |       |            |      |        |     |        |       |
| GO; GO:000568;    | 1       | F:proteolysis and peptidolysis; IEA.  |       |            |      |        |     |        |       |
| GO; GO:000550;    | 1       | F:transcription; IEA.   |       |            |      |        |     |        |       |
| GO; GO:0019079;   | 1       | P:viral genome replication; IEA.  |       |            |      |        |     |        |       |
| GO; GO:0019077;   | 1       | P:viral transformation; IEA.  |       |            |      |        |     |        |       |
| DR                | 1       | InterPro; IPR00032; Cys_Ser_trypsin.  |       |            |      |        |     |        |       |
| DR                | 1       | InterPro; IPR001410; DEAD.  |       |            |      |        |     |        |       |
| DR                | 1       | InterPro; IPR002522; HCV_capsid.  |       |            |      |        |     |        |       |
| DR                | 1       | InterPro; IPR00321; HCV_core.   |       |            |      |        |     |        |       |
| DR                | 1       | InterPro; IPR003519; HCV_env.   |       |            |      |        |     |        |       |
| DR                | 1       | InterPro; IPR002531; HCV_Ns1.   |       |            |      |        |     |        |       |
| DR                | 1       | InterPro; IPR002518; HCV_Ns2.   |       |            |      |        |     |        |       |
| DR                | 1       | InterPro; IPR000745; HCV_Ns4a.  |       |            |      |        |     |        |       |
| DR                | 1       | InterPro; IPR001490; HCV_Ns4b.  |       |            |      |        |     |        |       |
| DR                | 1       | InterPro; IPR002868; HCV_Ns5a.  |       |            |      |        |     |        |       |
| DR                | 1       | InterPro; IPR002516; HCV_RdRp.  |       |            |      |        |     |        |       |
| DR                | 1       | InterPro; IPR001650; Helicase_C.  |       |            |      |        |     |        |       |
| DR                | 1       | InterPro; IPR004109; Peptidase_C29.   |       |            |      |        |     |        |       |
| DR                | 1       | InterPro; IPR007095; RNA_Pol_DG_Ps.   |       |            |      |        |     |        |       |
| DR                | 1       | InterPro; IPR007094; RNA_Pol_PvVir.   |       |            |      |        |     |        |       |
| PFAM; PF01543;    | 1       | HCV_capsid; 1.  |       |            |      |        |     |        |       |
| PFAM; PF01542;    | 1       | HCV_core; 1.  |       |            |      |        |     |        |       |
| PFAM; PF01539;    | 1       | HCV_env; 1.   |       |            |      |        |     |        |       |
| PFAM; PF01560;    | 1       | HCV_Ns1; 1.   |       |            |      |        |     |        |       |
| PFAM; PF01538;    | 1       | HCV_Ns2; 1.   |       |            |      |        |     |        |       |
| PFAM; PF02907;    | 1       | HCV_Ns3; 1.   |       |            |      |        |     |        |       |
| PFAM; PF01546;    | 1       | HCV_Ns4a; 1.  |       |            |      |        |     |        |       |
| PFAM; PF01542;    | 1       | HCV_Ns4b; 1.  |       |            |      |        |     |        |       |
| PFAM; PF0106;     | 1       | HCV_Ns5a; 1.  |       |            |      |        |     |        |       |
| PFAM; PF00271;    | 1       | helicase_C; 1.  |       |            |      |        |     |        |       |
| PFAM; PF00098;    | 1       | viral_RdRp; 1.  |       |            |      |        |     |        |       |
| PRODOM; PD186062; | 1       | HCV_Ns1; 1.   |       |            |      |        |     |        |       |
| DR                | 1       | SMART; SM00487; DEXDC; 1.   |       |            |      |        |     |        |       |
| KW                | 1       | Coat protein; Envelope protein; Glycoprotein; Nonstructural protein; Polypeptide; RNA-directed RNA polymerase; Transferase; Transmembrane.                            |       |            |      |        |     |        |       |
| SEQUENCE          | 1       | SEQUENCE_3010_RA; 327023_MM; E075BD9CFD8D1261_CRC64;  |       |            |      |        |     |        |       |
| SO                | 1       | Query Match   | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent identity  | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent mismatch  | 4.7%  | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent gap   | 0.0%  | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent conservative  | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent similar   | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent dissimilar  | 4.7%  | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent identical   | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned   | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (conservative)  | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (similar)   | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (dissimilar)  | 4.7%  | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (identical)   | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (aligned)   | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (conservative, aligned)   | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (similar, aligned)  | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (dissimilar, aligned)   | 4.7%  | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (identical, aligned)  | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (aligned, aligned)  | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (conservative, aligned, aligned)  | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (similar, aligned, aligned)   | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (dissimilar, aligned, aligned)  | 4.7%  | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (identical, aligned, aligned)   | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (aligned, aligned, aligned)   | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (conservative, aligned, aligned, aligned)   | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (similar, aligned, aligned, aligned)  | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (dissimilar, aligned, aligned, aligned)   | 4.7%  | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (identical, aligned, aligned, aligned)  | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (aligned, aligned, aligned, aligned)  | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (conservative, aligned, aligned, aligned, aligned)  | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (similar, aligned, aligned, aligned, aligned)   | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (dissimilar, aligned, aligned, aligned, aligned)  | 4.7%  | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (identical, aligned, aligned, aligned, aligned)   | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (aligned, aligned, aligned, aligned, aligned)   | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (conservative, aligned, aligned, aligned, aligned, aligned)   | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (similar, aligned, aligned, aligned, aligned, aligned)  | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (dissimilar, aligned, aligned, aligned, aligned, aligned)   | 4.7%  | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (identical, aligned, aligned, aligned, aligned, aligned)  | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (aligned, aligned, aligned, aligned, aligned, aligned)  | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (conservative, aligned, aligned, aligned, aligned, aligned, aligned)  | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (similar, aligned, aligned, aligned, aligned, aligned, aligned)   | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (dissimilar, aligned, aligned, aligned, aligned, aligned, aligned)  | 4.7%  | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (identical, aligned, aligned, aligned, aligned, aligned, aligned)   | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (aligned, aligned, aligned, aligned, aligned, aligned, aligned)   | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (conservative, aligned, aligned, aligned, aligned, aligned, aligned, aligned)   | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (similar, aligned, aligned, aligned, aligned, aligned, aligned, aligned)  | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (dissimilar, aligned, aligned, aligned, aligned, aligned, aligned, aligned)   | 4.7%  | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (identical, aligned, aligned, aligned, aligned, aligned, aligned, aligned)  | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (aligned, aligned, aligned, aligned, aligned, aligned, aligned, aligned)  | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (conservative, aligned, aligned, aligned, aligned, aligned, aligned, aligned, aligned)  | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (similar, aligned, aligned, aligned, aligned, aligned, aligned, aligned, aligned)   | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (dissimilar, aligned, aligned, aligned, aligned, aligned, aligned, aligned, aligned)  | 4.7%  | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (identical, aligned, aligned, aligned, aligned, aligned, aligned, aligned, aligned)   | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (aligned, aligned, aligned, aligned, aligned, aligned, aligned, aligned, aligned)   | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (conservative, aligned, aligned, aligned, aligned, aligned, aligned, aligned, aligned, aligned)   | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (similar, aligned, aligned, aligned, aligned, aligned, aligned, aligned, aligned, aligned)  | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (dissimilar, aligned, aligned, aligned, aligned, aligned, aligned, aligned, aligned, aligned)   | 4.7%  | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (identical, aligned, aligned, aligned, aligned, aligned, aligned, aligned, aligned, aligned)  | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (aligned, aligned, aligned, aligned, aligned, aligned, aligned, aligned, aligned, aligned)  | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (conservative, aligned, aligned, aligned, aligned, aligned, aligned, aligned, aligned, aligned, aligned)  | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (similar, aligned, aligned, aligned, aligned, aligned, aligned, aligned, aligned, aligned, aligned)   | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (dissimilar, aligned, aligned, aligned, aligned, aligned, aligned, aligned, aligned, aligned, aligned)  | 4.7%  | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (identical, aligned, aligned, aligned, aligned, aligned, aligned, aligned, aligned, aligned, aligned)   | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (aligned, aligned, aligned, aligned, aligned, aligned, aligned, aligned, aligned, aligned, aligned)   | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (conservative, aligned, aligned)                                     | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (similar, aligned, aligned)  | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (dissimilar, aligned, aligned)                                       | 4.7%  | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (identical, aligned, aligned)  | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (aligned, aligned, aligned)  | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (conservative, aligned, aligned)                            | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (similar, aligned, aligned)                                 | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (dissimilar, aligned, aligned)                              | 4.7%  | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (identical, aligned, aligned)                               | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (aligned, aligned, aligned)                                 | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (conservative, aligned, aligned)                   | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (similar, aligned, aligned)                        | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (dissimilar, aligned, aligned)                     | 4.7%  | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (identical, aligned, aligned)                      | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (aligned, aligned, aligned)                        | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (conservative, aligned, aligned)          | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (similar, aligned, aligned)               | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (dissimilar, aligned, aligned)            | 4.7%  | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (identical, aligned, aligned)             | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (aligned, aligned, aligned)               | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (conservative, aligned, aligned) | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (similar, aligned, aligned)      | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (dissimilar, aligned, aligned)   | 4.7%  | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                | 1       | Percent aligned (identical, aligned, aligned)    | 95.3% | Score      | 174; | DB     | 12; | Length | 3010; |
| SO                |         |   |       |            |      |        |     |        |       |



QY 1 AHLQWVPPPLNVRGGRDAILTLCAVHPLIFDITKULLAATGQPLMVHQAGITKVVFVR 60  
 Db ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : |||||  
 AhLQWVPPPLNVRGGRDAILTLCAVHPLIFDITKULLAATGQPLMVHQAGITKVVFVR 914  
 QY 61 AQLIIRACMLVRKGAGGHVQMAFMKLALAITGTVYCHLTPQDWAHAGLRILAVEPV 120  
 Db ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : |||||  
 AQLIIRACMLVRKGAGGHVQMAFMKLALAITGTVYCHLTPQDWAHAGLRILAVEPV 974  
 QY 915 AQLIIRACMLVRKGAGGHVQMAFMKLALAITGTVYCHLTPQDWAHAGLRILAVEPV 120  
 Db ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : |||||  
 VFSIDMETKLTWGAADTAACGDITLGQPSARSGKEELIGPALSIEGRWRILAPITASQ 1034  
 QY 121 FESMEVKLTWGAADTAACGDITLGQPSARSGKEELIGPALSIEGRWRILAPITASQ 180  
 Db ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : |||||  
 VFSIDMETKLTWGAADTAACGDITLGQPSARSGKEELIGPALSIEGRWRILAPITASQ 1034  
 QY 181 QTRGLGCITSLGRDKNQVEGVQVSTATOSFLATCVNSYCWTFFHAGSKTLAGPK 240  
 Db ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : |||||  
 QTRGLGCITSLGRDKNQVEGVQVSTATOSFLATCVNSYCWTFFHAGSKTLAGPK 1094  
 QY 241 GPTIQMYINVDQDIVGVGWOAPPGARSMTCTCGSSDLIVTRHADVIPRRGDSRGSLLS 300  
 Db ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : |||||  
 1095 GPTIQMYINVDQDIVGVGWOAPPGARSMTCTCGSSDLIVTRHADVIPRRGDSRGSLLS 1154  
 QY 301 PRPVSYLKSSGGILCPSGHAVGIFRAVCTEGVAKADFIVVESMETTMR 352  
 Db ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : |||||  
 1155 PRPVSYLKSSGGILCPSGHAVGIFRAVCTEGVAKADFIVVESMETTMR 1206

Search completed: May 6, 2004, 09:35:48  
 Job time : 33.9219 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

M protein - protein search, using sw model

run on: May 6, 2004, 09:25:16 ; Search time 13.8154 Seconds  
(without alignments)  
1315.364 Million cell updates/sec

title: US-10-650-585-13  
exact score: 1842  
sequence: AHLQWIPPLNVRGGRDAIL.....RGWAKAVDFIPVSMETTMR 352

coring table: BloSUM62  
Gapop 10.0 , Gapext 0.5

searched: 389414 seqs, 51625971 residues

total number of hits satisfying chosen parameters: 389414

minimum DB seq length: 0  
maximum DB seq length: 200000000

ost-processing: Minimum Match 0%  
Maximum Match 100%  
listing first 45 summaries

Issued Patents AA:  
1: /cgn2\_6/prodata/2/iaa/5A\_COMB\_PEP: \*  
2: /cgn2\_6/prodata/2/iaa/5B\_COMB\_PEP: \*  
3: /cgn2\_6/prodata/2/iaa/6A\_COMB\_PEP: \*  
4: /cgn2\_6/prodata/2/iaa/6B\_COMB\_PEP: \*  
5: /cgn2\_6/prodata/2/iaa/PCRTUS\_COMB\_PEP: \*  
6: /cgn2\_6/prodata/2/iaa/backfile1.PEP: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| result No. | Score | Query | Match Length | DB ID | Description       |
|------------|-------|-------|--------------|-------|-------------------|
| 1          | 1766  | 95.9  | 2201         | 4     | US-09-539-601-6   |
| 2          | 1766  | 95.9  | 2001         | 4     | US-09-539-601-15  |
| 3          | 1766  | 95.9  | 3010         | 4     | US-09-539-601-3   |
| 4          | 1766  | 95.9  | 3010         | 4     | US-09-539-601-21  |
| 5          | 1766  | 95.9  | 3010         | 4     | US-09-539-601-27  |
| 6          | 1759  | 95.9  | 3010         | 4     | US-09-539-601-33  |
| 7          | 1757  | 95.4  | 1692         | 3     | US-09-263-933-4   |
| 8          | 1757  | 95.4  | 1692         | 4     | US-09-919-901-4   |
| 9          | 1757  | 95.4  | 2307         | 3     | US-09-263-933-2   |
| 10         | 1757  | 95.4  | 2307         | 4     | US-09-919-901-2   |
| 11         | 1754  | 95.2  | 1692         | 3     | US-09-263-933-11  |
| 12         | 1754  | 95.2  | 1692         | 4     | US-09-919-901-11  |
| 13         | 1754  | 95.2  | 2307         | 3     | US-09-263-933-9   |
| 14         | 1754  | 95.2  | 2307         | 4     | US-09-919-901-9   |
| 15         | 1748  | 94.9  | 3010         | 3     | US-09-014-416-3   |
| 16         | 1745  | 94.7  | 1692         | 3     | US-09-263-933-18  |
| 17         | 1745  | 94.7  | 1692         | 4     | US-09-919-901-18  |
| 18         | 1745  | 94.7  | 2307         | 3     | US-09-263-933-16  |
| 19         | 1745  | 94.7  | 2307         | 4     | US-09-919-901-16  |
| 20         | 1699  | 92.2  | 2013         | 1     | US-08-324-977-12  |
| 21         | 1699  | 92.2  | 2013         | 2     | US-08-384-616-12  |
| 22         | 1699  | 92.2  | 2013         | 2     | US-08-904-686A-12 |
| 23         | 1699  | 92.2  | 2013         | 3     | US-08-315-850-12  |
| 24         | 1699  | 92.2  | 2201         | 4     | US-08-352-981A-2  |
| 25         | 1699  | 92.2  | 2620         | 1     | US-08-324-977-32  |
| 26         | 1699  | 92.2  | 2620         | 1     | US-08-616-32      |
| 27         | 1699  | 92.2  | 2620         | 2     | US-08-904-686A-32 |

RESULT 2

RESULT 1

US-09-539-601-6

; Sequence 6, Application US/09539601C  
; Parent No. 6530343  
; GENERAL INFORMATION:  
; APPLICANT: BartterschLAGER, Ralf FW  
; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System  
; FILE REFERENCE: all sequences  
; CURRENT APPLICATION NUMBER: US/09-539, 601C  
; CURRENT FILING DATE: 2001-08-30  
; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY  
; EARLIER FILING DATE: 1999-04-03  
; NUMBER OF SEQ ID NOS: 51  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 2201  
; TYPE: PRT  
; ORGANISM: Hepatitis C virus  
; US-09-539-601-6

Query Match 95.9%; Score 1766; DB 4; Length 2201;  
Best Local Similarity 94.3%; Pred. No. 1.2e-170;  
Matches 332; Conservativeness 11; Mismatches 9; Indels 0; Gaps 0;

QY 1 AHLQWIPPLNVRGGRDAILLTCAVHPPELFDITKLLAIFGLPVLNQAGITKPYFVR 60  
Db 46 AHLQWIPPLNVRGGRDAILLTCAVHPPELFTIKLLAIFGLPVLNQAGITKPYFVR 105  
QY 61 AQLGLTRACMLVRKAGHYTOMAFMKLAALTGTYYDHPLTDQWDWAHAGLDLAVAPV 120  
Db 106 AQLGLTRACMLVRKAGHYVQVALMKLAALTGTYYDHPLTDQWDWAHAGLDLAVAPV 165  
QY 121 IFSMDMEVKILTWGADTAACGDIISLGIPVPSARRGRELTLGGPADNFGQGWRLLAPIAYSQ 180  
Db 166 VFSMDMERTKVITWGADTAACGDIILGLPVPSARRGRELTHLGPADSLEGQGWRLAPIAYSQ 225  
QY 181 QTRGLIICITSLTGDKNOVEGEVQWSATQSFLATCTCNGVWVFHGSKTLAGPK 240  
Db 226 QTRGLIICILTSLGDRDNQVEGEVQWSATQSFLATCTCNGVWVFHGSKTLAGPK 285  
QY 241 GPTIQTOMTNVQDLDVWQAPPGARSAWTPCTGSSDIYLVTRHADIVPVRREGDGSLSLS 300  
Db 286 GPTIQTOMTNVQDLDVWQAPPGARSAWTPCTGSSDIYLVTRHADIVPVRREGDGSLSLS 345  
QY 301 PRPVSTLKGSSGGPLCPGSHAVGIGRTRGVAKAVFIPVSMETTMR 352  
Db 346 PRPVSTLKGSSGGPLCPGSHAVGIGRTRGVAKAVFIPVSMETTMR 397

ALIGNMENTS

US-09-539-601-15  
; Sequence 15, Application US/09539601C  
; Patent No. 663043  
; GENERAL INFORMATION:  
; APPLICANT: Bartenschlager, Ralf FW  
; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System  
; FILE REFERENCE: all sequences  
; CURRENT APPLICATION NUMBER: US/09/539,601C  
; CURRENT FILING DATE: 2001-08-30  
; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY  
; EARLIER FILING DATE: 1999-04-03  
; NUMBER OF SEQ ID NOS: 51  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO: 15  
; LENGTH: 2201  
; TYPE: PRT  
; ORGANISM: Hepatitis C virus  
; US-09-539-601-15

Query Match 95.9%; Score 1766; DB 4; Length 2201;  
Best Local Similarity 94.3%; Pred. No. 1.2e-170;  
Matches 332; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

```
QY 1 AHLQWIPPIVNRGGRDAILTLCAVHPPEIIFTKLIAIFGMLMVQAGITKPYFVR 60
Db 46 AHLQWIPPIVNRGGRDAILTLCAVHPPEIIFTKLIAIFGMLMVQAGITKPYFVR 105
```

QY 61 AQLIRACMVRKAAGGHYOMAFMKLAALTGTYYDHILPQDWAHGLRDLAVAEPV 120
Db 106 AQLIRACMVLVRKVAGHYOMALMMLAALTGTYYDHILPQDWAHGLRDLAVAEPV 165

QY 121 IFSDMEVKITWGAATAACDDISGLPVSAARRGREBILGPADNPEGQMLVQAGITKPYFVR 180
Db 166 VFSDEMETKVITWGAATAACDDISGLPVSAARRGREBILGPADNPEGQMLVQAGITKPYFVR 225

QY 181 QTRGLIGCITSLSLGRDKNQVEGEQVNVSTATOFLATCNGVWTIVHGAGSKTLAPK 240
Db 226 QTRGLIGCITSLSLGRDKNQVEGEQVNVSTATOFLATCNGVWTIVHGAGSKTLAPK 285

QY 241 GPITQMTNTQDQDVGWQAPPGRASMTCTCGSSDLVYLTTRHADVIPRRGDSRGSLIS 300
Db 286 GPITQMTNTQDQDVGWQAPPGRASLTCTCGSSDLVYLTTRHADVIPRRGDSRGSLIS 345

QY 301 PRPVSYLKGSGGPILCPSPGHAVGIFRAAVCTRGVAKAVDFPVVESMETMR 352
Db 346 PRPVSYLKGSGGPILCPSPGHAVGIFRAAVCTRGVAKAVDFPVVESMETMR 397

RESULT 3  
US-09-539-601-3  
; Sequence 3, Application US/09539601C  
; Patent No. 663043  
; GENERAL INFORMATION:  
; APPLICANT: Bartenschlager, Ralf FW  
; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System  
; FILE REFERENCE: all sequences  
; CURRENT APPLICATION NUMBER: US/09/539,601C  
; CURRENT FILING DATE: 2001-08-30  
; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY  
; EARLIER FILING DATE: 1999-04-03  
; NUMBER OF SEQ ID NOS: 51  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO: 21  
; LENGTH: 3010  
; TYPE: PRT  
; ORGANISM: Hepatitis C virus  
; US-09-539-601-21

Query Match 95.9%; Score 1766; DB 4; Length 3010;  
Best Local Similarity 94.3%; Pred. No. 1.3e-170;  
Matches 332; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

```
QY 1 AHLQWIPPIVNRGGRDAILTLCAVHPPEIIFTKLIAIFGMLMVQAGITKPYFVR 60
Db 855 AHLQWIPPIVNRGGRDAILTLCAVHPPEIIFTKLIAIFGMLMVQAGITKPYFVR 914
```

QY 61 AQLIRACMVRKAAGGHYOMAFMKLAALTGTYYDHILPQDWAHGLRDLAVAEPV 120
Db 915 AQLIRACMVLVRKVAGHYOMALMMLAALTGTYYDHILPQDWAHGLRDLAVAEPV 974

QY 121 IFSDMEVKITWGAATAACDDISGLPVSAARRGREBILGPADNPEGQMLVQAGITKPYFVR 180
Db 975 VFSDEMETKVITWGAATAACDDISGLPVSAARRGREBILGPADNPEGQMLVQAGITKPYFVR 1034

QY 181 QTRGLIGCITSLSLGRDKNQVEGEQVNVSTATOFLATCNGVWTIVHGAGSKTLAPK 240
Db 1035 QTRGLIGCITSLSLGRDKNQVEGEQVNVSTATOFLATCNGVWTIVHGAGSKTLAPK 1094

QY 241 GPITQMTNTQDQDVGWQAPPGRASMTCTCGSSDLVYLTTRHADVIPRRGDSRGSLIS 300
Db 1095 GPITQMTNTQDQDVGWQAPPGRASLTCTCGSSDLVYLTTRHADVIPRRGDSRGSLIS 1154

QY 301 PRPVSYLKGSGGPILCPSPGHAVGIFRAAVCTRGVAKAVDFPVVESMETMR 352
Db 1155 PRPVSYLKGSGGPILCPSPGHAVGIFRAAVCTRGVAKAVDFPVVESMETMR 1206

RESULT 5  
US-09-539-601-27  
; Sequence 27, Application US/09539601C  
; Patent No. 6630343  
; GENERAL INFORMATION:  
; APPLICANT: Bartenschlager, Ralf FW  
; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System  
; FILE REFERENCE: all sequences  
; CURRENT APPLICATION NUMBER: US/09/539,601C  
; CURRENT FILING DATE: 2001-08-30  
; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY  
; EARLIER FILING DATE: 1999-04-03  
; NUMBER OF SEQ ID NOS: 51  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 27  
; LENGTH: 3010  
; TYPE: PRT  
; ORGANISM: Hepatitis C virus  
; US-09-539-601-27

Query Match 95.9%; Score 1766; DB 4; Length 3010;  
Best Local Similarity 94.3%; Pred. No. 1-9e-170; Mismatches 9; Indels 0; Gaps 0;  
Matches 332; Conservative 11; Predicted 1-9e-170;

QY 1 AHLQWIPPLNVRKGAGHVVOMAFMKAALGTGTYVHDHTPLQDWAHAGRLAVAPEV 60  
Db 855 AHLQWIPPLNVRKGAGHVVOMAFMKAALGTGTYVHDHTPLQDWAHAGRLAVAPEV 914

QY 61 AQLIRACMLVRKAAGHVVOMAFMKAALGTGTYVHDHTPLQDWAHAGRLAVAPEV 120  
Db 915 AQLIRACMLVRKAAGHVVOMAFMKAALGTGTYVHDHTPLQDWAHAGRLAVAPEV 974

QY 121 IFSDMETKIIWGAIDTAAAGDTISGLPVPSARGREBLLGPADNFECQWGRLLAPTA 180  
Db 975 VFSDEMKITWGAIDTAAAGDTISGLPVPSARGREBLLGPADNFECQWGRLLAPTA 180

QY 181 QTRGLIGCILSITLGDRKQVEGEVQVSTATQSFATCNGVCMVVFHAGSKTLAGPK 240  
Db 1095 GPITQMYTNVDLQWQAPPGRSLTPCTCGSSDLVLYLVRHADVIPRRGDSRSLS 1154

QY 301 PRPVSYLKGSQQGLCPSPGHAVGIFRAVCTRGVAKAVDPVSMETMR 352  
Db 1155 PRPVSYLKGSQQGLCPSPGHAVGIFRAVCTRGVAKAVDPVSMETMR 1206

PRPVSYLKGSQQGLCPSPGHAVGIFRAVCTRGVAKAVDPVSMETMR 1206

RESULT 7  
US-09-263-933-4  
; Sequence 4, Application US/09263933  
; Patent No. 6280940  
; GENERAL INFORMATION:  
; APPLICANT: Potts, Karen E.  
; APPLICANT: Jackson, Roberta L.  
; APPLICANT: Patrick, Amy K.  
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE  
; FILE REFERENCE: 0125-0005A  
; CURRENT APPLICATION NUMBER: US/09/263, 933  
; CURRENT FILING DATE: 1999-03-08  
; EARLIER APPLICATION NUMBER: 1998-08-05  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 1692  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; US-09-263-933-4

Query Match 95.4%; Score 1757; DB 3; Length 1692;  
Best Local Similarity 93.8%; Pred. No. 6.7e-170; Mismatches 10; Indels 0; Gaps 0;  
Matches 330; Conservative 12; Predicted 6.7e-170;

QY 1 AHLQWIPPLNVRKGAGHVVOMAFMKAALGTGTYVHDHTPLQDWAHAGRLAVAPEV 60  
Db 134 AHLWIPPLNVRKGAGHVVOMAFMKAALGTGTYVHDHTPLQDWAHAGRLAVAPEV 193

QY 61 AQLIRACMLVRKAAGHVVOMAFMKAALGTGTYVHDHTPLQDWAHAGRLAVAPEV 120  
Db 194 AQLIRACMLVRKAAGHVVOMAFMKAALGTGTYVHDHTPLQDWAHAGRLAVAPEV 253

QY 121 IFSDMETKIIWGAIDTAAAGDTISGLPVPSARGREBLLGPADNFECQWGRLLAPTA 180  
Db 254 VFSDEMKITWGAIDTAAAGDTISGLPVPSARGREBLLGPADNFECQWGRLLAPTA 313

QY 181 QTRGLIGCILSITLGDRKQVEGEVQVSTATQSFATCNGVCMVVFHAGSKTLAGPK 240  
Db 314 QTRGLIGCILSITLGDRKQVEGEVQVSTATQSFATCNGVCMVVFHAGSKTLAGPK 373

QY 241 GPITQMYTNVDLQWQAPPGRSLTPCTCGSSDLVLYLVRHADVIPRRGDSRSLS 300  
Db 374 GPITQMYTNVDLQWQAPPGRSLTPCTCGSSDLVLYLVRHADVIPRRGDSRSLS 433

QY 301 PRPVSYLKGSQQGLCPSPGHAVGIFRAVCTRGVAKAVDPVSMETMR 352

RESULT 8  
US-09-919-901-4  
Sequence 4, Application US/09919901  
; Patent No. 6599738  
; GENERAL INFORMATION:  
; APPLICANT: Potts, Karen E.  
; APPLICANT: Jackson, Roberta L.  
; APPLICANT: Patrick, Amy K.  
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE  
; FILE REFERENCE: 0125-0005A  
; CURRENT APPLICATION NUMBER: US/09/919-901  
; CURRENT FILING DATE: 2001-08-02  
; PRIORITY APPLICATION NUMBER: 09/263-933  
; PRIORITY FILING DATE: 1999-02-08  
; PRIORITY APPLICATION NUMBER: 09/129, 611  
; PRIORITY FILING DATE: 1998-08-05  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 2307  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; US-09-263-933-2  
; Query Match 95.4%; Score 1757; DB 3; Length 2307;  
; Best Local Similarity 93.8%; Pred. No. 1.e-169; Indels 0; Gaps 0;  
; Matches 330; Conservative 12; Mismatches 10; Indels 0; Gaps 0;  
Qy 1 AHLQWIPPLNRRGRDAILLTCAHPPELIDTICKLLAIFGPMIYQAGITKVPYFVR 60  
Db 226 AHLHWIPPLNRRGRDAILLTCAHPPELIDTICKLLAIFGPMIYQAGITKVPYFVR 285  
Qy 61 AHLQWIPPLNRRGRDAILLTCAHPPELIDTICKLLAIFGPMIYQAGITKVPYFVR 120  
Db 286 AHLQWIPPLNRRGRDAILLTCAHPPELIDTICKLLAIFGPMIYQAGITKVPYFVR 345  
Qy 121 IRSDEMEKITTWDATACGDTISGLPVSAARGRETLIGPADNFEGOGWRILAPITYSQ 180  
Db 346 VPSDEMEKITTWDATACGDTISGLPVSAARGRETLIGPADNFEGOGWRILAPITYSQ 405  
Qy 181 QTRGLLGCCITSLGRDKNQVGEVQVSTATOSFLATCVNCVWTVFHGAGSKTLAGPK 240  
Db 466 GPITQMYTNVQDVLGVQAPGARSLSPTCQSSDLYLVTRHADVIPVRRGDSRGSLLS 525  
Qy 301 PRPVSYLKGGSGGPLCPSGHAVGIRRAVCTRGVAKADVIPVVEMETMR 352  
Db 526 PRPVSYLKGGSGGPLCPSGHAVGIRRAVCTRGVAKADVIPVVEMETMR 577

Query Match 95.4%; Score 1757; DB 4; Length 1692;  
Best Local Similarity 93.8%; Pred. No. 6.7e-170; Mismatches 10; Indels 0; Gaps 0;  
Matches 330; Conservative 12; Mismatches 10; Indels 0; Gaps 0;  
OTHER INFORMATION: :  
; US-09-919-901-4  
Qy 1 AHLQWIPPLNRRGRDAILLTCAHPPELIDTICKLLAIFGPMIYQAGITKVPYFVR 60  
Db 134 AHLHWIPPLNRRGRDAILLTCAHPPELIDTICKLLAIFGPMIYQAGITKVPYFVR 193  
Qy 61 AHLQWIPPLNRRGRDAILLTCAHPPELIDTICKLLAIFGPMIYQAGITKVPYFVR 120  
Db 194 AHLQWIPPLNRRGRDAILLTCAHPPELIDTICKLLAIFGPMIYQAGITKVPYFVR 253  
Qy 121 IRSDEMEKITTWDATACGDTISGLPVSAARGRETLIGPADNFEGOGWRILAPITYSQ 180  
Db 254 VPSDEMEKITTWDATACGDTISGLPVSAARGRETLIGPADNFEGOGWRILAPITYSQ 313  
Qy 181 QTRGLLGCCITSLGRDKNQVGEVQVSTATOSFLATCVNCVWTVFHGAGSKTLAGPK 240  
Db 314 QTRGLLGCCITSLGRDKNQVGEVQVSTATOSFLATCVNCVWTVFHGAGSKTLAGPK 373  
Qy 241 GPITQMYTNVQDVLGVQAPGARSLSPTCQSSDLYLVTRHADVIPVRRGDSRGSLLS 300  
Db 374 GPITQMYTNVQDVLGVQAPGARSLSPTCQSSDLYLVTRHADVIPVRRGDSRGSLLS 433  
Qy 301 PRPVSYLKGGSGGPLCPSGHAVGIRRAVCTRGVAKADVIPVVEMETMR 352  
Db 434 PRPVSYLKGGSGGPLCPSGHAVGIRRAVCTRGVAKADVIPVVEMETMR 485

RESULT 9  
US-09-263-933-2  
; Sequence 2, Application US/09263933  
; Patent No. 6280940  
; GENERAL INFORMATION:  
; APPLICANT: Potts, Karen E.  
; APPLICANT: Jackson, Roberta L.  
; APPLICANT: Patrick, Amy K.  
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE  
; FILE REFERENCE: 0125-0005A  
; CURRENT APPLICATION NUMBER: US/09/919-901  
; CURRENT FILING DATE: 2001-08-02  
; PRIORITY APPLICATION NUMBER: 09/263-933  
; PRIORITY FILING DATE: 1999-02-08  
; PRIORITY APPLICATION NUMBER: 09/129, 611  
; PRIORITY FILING DATE: 1998-08-05  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 2307  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; US-09-919-901-2  
; Query Match 95.4%; Score 1757; DB 4; Length 2307;  
; Best Local Similarity 93.8%; Pred. No. 1.e-169; Indels 0; Gaps 0;  
; Matches 330; Conservative 12; Mismatches 10; Indels 0; Gaps 0;  
Qy 1 AHLQWIPPLNRRGRDAILLTCAHPPELIDTICKLLAIFGPMIYQAGITKVPYFVR 60  
Db 226 AHLHWIPPLNRRGRDAILLTCAHPPELIDTICKLLAIFGPMIYQAGITKVPYFVR 285

RESULT 11  
 US-09-263-933-11  
 ; Sequence 11, Application US/09263933  
 ; Patent No. 6280940  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Potts, Karen E.  
 ; APPLICANT: Jackson, Roberta L.  
 ; APPLICANT: Patick, Amy K.  
 ; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
 ; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE  
 ; FILE REFERENCE: 0125-0005A  
 ; CURRENT APPLICATION NUMBER: US/09/263, 933  
 ; CURRENT FILING DATE: 1999-03-08  
 ; EARLIER APPLICATION NUMBER: 09/129, 611  
 ; EARLIER FILING DATE: 1998-08-05  
 ; NUMBER OF SEQ ID NOS: 33  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO: 11  
 ; LENGTH: 1692  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE: :  
 ; OTHER INFORMATION: :  
 US-09-919-901-11  
 ; Query Match 95.2%; Score 1754; DB 4; Length 1692;  
 ; Best Local Similarity 93.5%; Pred. No. 1.4e-169; Matches 329; Conservative 13; Mismatches 10; Indels 0; Gaps 0;  
 ; Matches 329; Conservative 13; Mismatches 10; Indels 0; Gaps 0;  
 QY 1 AHIQWIPPLNVRGGRDAITLTCAVPHPELFDITKLLAIFPMLVQAGCTKVPYFVR 60  
 Db 134 AHLHWIPPLNVRGGRDAITLIMCAVPHPELFDITKLLAIFPMLVQAGCTKVPYFVR 193  
 QY 61 AOGLIIRACMLVRKAAGHHVQMAFMKLAALTGTIVYDHITPLQDWAHGLRDLAVAPEV 120  
 Db 194 AOGLIIRACMLVRKAAGHHVQMAFMKLAALTGTIVYDHITPLQDWAHGLRDLAVAPEV 253  
 QY 121 IFSDMEVKITIWGADTAACGDIISGLPVASRREBILGPADNFEGQWRLAPIYSQ 180  
 Db 314 QTRGLLGCCITSLSLTDQGRDKNVQEVGVQVSTATOSFLATCUNGVCWTFHGASKTLAGPK 373  
 QY 241 GPITOMTYNTWDQDLVGWQAPPGRASMTPTCTGSSDLYLVLTRHADVIPVRRGDSRSGLS 300  
 Db 374 GPITOMTYNTWDQDLVGWQAPPGRASLTPTCTGSSDLYLVLTRHADVIPVRRGDSRSGLS 433  
 QY 301 PRPSYLIKSSGGPLCPGHAVGIFRAVCTGVAKAVDFPVESMETMR 352  
 Db 434 PRPVSYLKGSAGGPLCPGHAVGIFRAVCTGVAKAVDFPVESMETMR 485  
 RESULT 12  
 US-09-919-901-11  
 ; Sequence 11, Application US/09263933  
 ; Patent No. 6280940  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Potts, Karen E.  
 ; APPLICANT: Jackson, Roberta L.  
 ; APPLICANT: Patick, Amy K.  
 ; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
 ; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE  
 ; FILE REFERENCE: 0125-0005A  
 ; CURRENT APPLICATION NUMBER: US/09/919, 901  
 ; CURRENT FILING DATE: 2001-08-02  
 ; PRIOR APPLICATION NUMBER: 09/263, 933  
 ; PRIOR APPLICATION NUMBER: 09/129, 611  
 ; PRIOR FILING DATE: 1998-08-05  
 ; NUMBER OF SEQ ID NOS: 33  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO: 11  
 ; LENGTH: 1692  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE: :  
 ; OTHER INFORMATION: :  
 US-09-919-901-11  
 ; Query Match 95.2%; Score 1754; DB 4; Length 1692;  
 ; Best Local Similarity 93.5%; Pred. No. 1.4e-169; Matches 329; Conservative 13; Mismatches 10; Indels 0; Gaps 0;  
 ; Matches 329; Conservative 13; Mismatches 10; Indels 0; Gaps 0;  
 QY 1 AHIQWIPPLNVRGGRDAITLTCAVPHPELFDITKLLAIFPMLVQAGCTKVPYFVR 60  
 Db 134 AHLHWIPPLNVRGGRDAITLIMCAVPHPELFDITKLLAIFPMLVQAGCTKVPYFVR 193  
 QY 61 AOGLIIRACMLVRKAAGHHVQMAFMKLAALTGTIVYDHITPLQDWAHGLRDLAVAPEV 120  
 Db 194 AOGLIIRACMLVRKAAGHHVQMAFMKLAALTGTIVYDHITPLQDWAHGLRDLAVAPEV 253  
 QY 121 IFSDMEVKITIWGADTAACGDIISGLPVASRREBILGPADNFEGQWRLAPIYSQ 180  
 Db 314 QTRGLLGCCITSLSLTDQGRDKNVQEVGVQVSTATOSFLATCUNGVCWTFHGASKTLAGPK 373  
 QY 241 GPITOMTYNTWDQDLVGWQAPPGRASMTPTCTGSSDLYLVLTRHADVIPVRRGDSRSGLS 300  
 Db 374 GPITOMTYNTWDQDLVGWQAPPGRASLTPTCTGSSDLYLVLTRHADVIPVRRGDSRSGLS 433  
 QY 301 PRPSYLIKSSGGPLCPGHAVGIFRAVCTGVAKAVDFPVESMETMR 352  
 Db 434 PRPVSYLKGSAGGPLCPGHAVGIFRAVCTGVAKAVDFPVESMETMR 485  
 RESULT 13  
 US-09-263-933-9  
 ; Sequence 9, Application US/09263933  
 ; Patent No. 6280940  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Potts, Karen E.  
 ; APPLICANT: Jackson, Roberta L.  
 ; APPLICANT: Patick, Amy K.  
 ; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
 ; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE  
 ; FILE REFERENCE: 0125-0005A  
 ; CURRENT APPLICATION NUMBER: US/09/263, 933  
 ; CURRENT FILING DATE: 1999-03-09  
 ; EARLIER APPLICATION NUMBER: 09/129, 611  
 ; EARLIER FILING DATE: 1998-08-05  
 ; NUMBER OF SEQ ID NOS: 33

; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO: 9  
; LENGTH: 2307  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; US-09-263-933-9  
Query Match 95.2%; Score 1754; DB 3; Length 2307;  
Best Local Similarity 93.5%; Pred. No. 2.1e-169; Matches 329; Conservative 13; Mismatches 10; Indels 0; Gaps 0;  
QY 1 AHLQWIPPLNVRGRDATTILTCVHPBLIFDITKULLAIFPMLVQAGITKVPFVR 60  
Db 226 AHLHWIPPLNVRGRDATTILTCVHPBLIFDITKULLAIFPMLVQAGITKVPFVR 285  
Db 286 AQGLIHACMLVRKVAGGHVYOMAFMQLGALTGTIVYHNLPLRDWAHAGRLDAVAVFPV 120  
QY 61 AQGLTRACMLVRKVAGGHVYOMAFMQLGALTGTIVYHNLPLRDWAHAGRLDAVAVFPV 120  
Db 286 AQGLIHACMLVRKVAGGHVYOMAFMQLGALTGTIVYHNLPLRDWAHAGRLDAVAVFPV 345  
QY 121 IFSDMEVKILTGWADTAACCDIISLPLVSARRGKEIFLGPLDSLEGQRWLRAPITAVSQ 180  
Db 346 VFSDMEVKILTGWADTAACCDIISLPLVSARRGKEIFLGPLDSLEGQRWLRAPITAVSQ 405  
QY 181 QTRGLGICITSLSLGRDKNOVEGEVQVSTATOSFLATCUNGVCWTYHGAGSKTLAGPK 240  
Db 466 GPITQMTNTDODLGWQPGARSMTPCCGSSDLVYTRHADVIPVRGRGSRSGLS 300  
QY 301 PRPVSYLKGSGGPGLCPSCSHAVGIFRAVCTRGVAKAVDFPVVESMETMR 352  
Db 526 PRPVSYLKGSAGGPUCPSCHAVGIFRAVCTRGVAKAVDFPVVESMETMR 577  
RESULT 14  
US-09-919-901-9  
Sequence 9, Application US/09919901  
; Patent No. 6599738  
GENERAL INFORMATION:  
APPLICANT: Potts, Karen E.  
APPLICANT: Jackson, Robertta L.  
APPLICANT: Patrick, Amy K.  
TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
FILE REFERENCE: 0125-005A  
CURRENT APPLICATION NUMBER: US/09/919, 901  
CURRENT FILING DATE: 2001-08-02  
PRIOR APPLICATION NUMBER: 09/263, 933  
PRIOR FILING DATE: 1999-02-08  
PRIOR APPLICATION NUMBER: 09/129, 611  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 9  
LENGTH: 3010  
TYPE: PRT  
ORGANISM: Hepatitis C virus  
; US-09-014-416-3  
Query Match 94.9%; Score 1748; DB 3; Length 3010;  
Best Local Similarity 92.9%; Pred. No. 1.3e-168; Matches 327; Conservative 14; Mismatches 11; Indels 0; Gaps 0;  
QY 1 AHLQWIPPLNVRGRDATTILTCVHPBLIFDITKULLAIFPMLVQAGITKVPFVR 60  
Db 855 AHMQWIPPLNVRGRDATTILTCVHPBLIFDITKULLAIFPMLVQAGITKVPFVR 914  
QY 61 AQGLRACMLVRKVAGGHVYOMAFMQLGALTGTIVYHNLPLRDWAHAGRLDAVAVFPV 120  
Db 915 AQGLRACMLVRKVAGGHVYOMAFMQLGALTGTIVYHNLPLRDWAHAGRLDAVAVFPV 974  
QY 121 IFSDMEVKILTGWADTAACCDIISLPLVSARRGKEIFLGPLDSLEGQRWLRAPITAVSQ 180  
Db 975 VFSAEMETKVITGWADTAACCDIISLPLVSARRGKEIFLGPLDSLEGQRWLRAPITAVSQ 1034  
QY 181 QTRGLGICITSLSLGRDKNOVEGEVQVSTATOSFLATCUNGVCWTYHGAGSKTLAGPK 240  
Db 1035 QTRGLGICITSLSLGRDKNOVEGEVQVSTATOSFLATCUNGVCWTYHGAGSKTLAGPK 1094  
QY 241 GPITQMTNTDODLGWQPGARSMTPCCGSSDLVYTRHADVIPVRGRGSRSGLS 300  
Db 1095 GPITQMTNTDODLGWQPGARSMTPCCGSSDLVYTRHADVIPVRGRGSRSGLS 1154  
QY 301 PRPVSYLKGSGGPGLCPSCSHAVGIFRAVCTRGVAKAVDFPVVESMETMR 352  
Db 1155 PRPVSYLKGSGGPGLCPSCSHAVGIFRAVCTRGVAKAVDFPVVESMETMR 1206  
QY 61 AQGLTRACMLVRKVAGGHVYOMAFMQLGALTGTIVYHNLPLRDWAHAGRLDAVAVFPV 120

Search completed: May 6, 2004, 09:39:04

Fri May 7 13:37:03 2004

Job time : 14.8154 secs

us-10-650-585-13.rai

Page 7





Db 181 QTRGLIGCITSLGRDKNQVEGVQVSTATOSFLATCVNGCWTVPFHGAGSKTLAGPK 240 ; PRIOR FILING DATE: 2000-12-15  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO: 12  
; LENGTH: 380  
; TYPE: PRT  
; ORGANISM: HCV  
; US-10-017-736-12

RESULT 2  
US-10-650-585-13  
; Sequence 13, Application US/10650585  
; Publication No. US20040077066A1  
; GENERAL INFORMATION:  
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.  
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease  
; FILE REFERENCE: 13/082  
; CURRENT APPLICATION NUMBER: US/10/650, 585  
; CURRENT FILING DATE: 2003-08-28  
; PRIOR APPLICATION NUMBER: US/10/017, 736A  
; PRIOR FILING DATE: 2001-12-14  
; PRIOR FILING DATE: 2000-12-15  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO: 13  
; LENGTH: 352  
; TYPE: PRT  
; ORGANISM: HCV  
; US-10-650-585-13

Query Match 100.0%; Score 1842; DB 16; Length 380;  
Best Local Similarity 100.0%; Pred. No. 5. 7e-10; Mismatches 0; Indels 0; Gaps 0;  
Matches 352; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AHLQWIPPLNVRGDRDAILITCAVHBLIFDITKLLIAIFGPLMLQAGITKVPYVR 60  
Db 29 AHLQWIPPLNVRGDRDAILITCAVHBLIFDITKLLIAIFGPLMLQAGITKVPYVR 88  
; 61 AQGLIRACMVRKAAGGHVQMAFMKLALITGTIVYDILPQDWAHGLRDLAVAPEV 120  
; 89 AQGLIRACMVRKAAGGHVQMAFMKLALITGTIVYDILPQDWAHGLRDLAVAPEV 148  
; 121 IFSDMEVKITTWGADTAACDITSGLPVSARRGREBILGPADNPEGQWRLLAPITAYSQ 180  
; 149 IFSDMEVKITTWGADTAACDITSGLPVSARRGREBILGPADNPEGQWRLLAPITAYSQ 208  
; 181 QTRGLIGCITSLGRDKNQVEGVQVSTATOSFLATCVNGCWTVPFHGAGSKTLAGPK 240  
; 209 QTRGLIGCITSLGRDKNQVEGVQVSTATOSFLATCVNGCWTVPFHGAGSKTLAGPK 268  
; 241 GPITQMYNTDQDVGWQPGARSMTPTCGSSDLYLVTREHADVIPRRGDSRGSILS 300  
; 269 GPITQMYNTDQDVGWQPGARSMTPTCGSSDLYLVTREHADVIPRRGDSRGSILS 328  
; 301 PRPVSYLKGSGGPPLCPSGHAVGIFRAAVCTRGAVAKAVDFIPVESMETTMR 380  
; 329 PRPVSYLKGSGGPPLCPSGHAVGIFRAAVCTRGAVAKAVDFIPVESMETTMR 380

RESULT 4  
US-10-650-585-12  
; Sequence 12, Application US/10650585  
; Publication No. US20040077066A1  
; GENERAL INFORMATION:  
; APPLICANT: Boehringer Ingelheim (Canada) Ltd.  
; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease  
; FILE REFERENCE: 13/082  
; CURRENT APPLICATION NUMBER: US/10/650, 585  
; CURRENT FILING DATE: 2003-08-28  
; PRIOR APPLICATION NUMBER: US/10/017, 736A  
; PRIOR FILING DATE: 2001-12-14  
; PRIOR APPLICATION NUMBER: 60/256, 031  
; PRIOR FILING DATE: 2000-12-15  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO: 12  
; LENGTH: 380  
; TYPE: PRT  
; ORGANISM: HCV  
; US-10-650-585-12

Query Match 100.0%; Score 1842; DB 16; Length 380;  
Best Local Similarity 100.0%; Pred. No. 6. 4e-10; Mismatches 0; Indels 0; Gaps 0;  
Matches 352; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AHLQWIPPLNVRGDRDAILITCAVHBLIFDITKLLIAIFGPLMLQAGITKVPYVR 60  
Db 29 AHLQWIPPLNVRGDRDAILITCAVHBLIFDITKLLIAIFGPLMLQAGITKVPYVR 88  
; 61 AQGLIRACMVRKAAGGHVQMAFMKLALITGTIVYDILPQDWAHGLRDLAVAPEV 120  
; 89 AQGLIRACMVRKAAGGHVQMAFMKLALITGTIVYDILPQDWAHGLRDLAVAPEV 148  
; 121 IFSDMEVKITTWGADTAACDITSGLPVSARRGREBILGPADNPEGQWRLLAPITAYSQ 180

RESULT 6

US-10-650-585-11

; Sequence 11; Application US/10050585

; Publication No. US2004007706A1

; GENERAL INFORMATION:

; APPLICANT: Boehringer Ingelheim (Canada) Ltd.

; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease

; FILE REFERENCE: 13/082

; CURRENT APPLICATION NUMBER: US/10/017,736

; CURRENT FILING DATE: 2001-12-14

; PRIORITY NUMBER: 60/256,031

; NUMBER OF SEQ ID NOS: 21

; SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO 11

; LENGTH: 393

; TYPE: PRT

; ORGANISM: HCV

US-10-650-585-11

Query Match 100.0%; Score 1842; DB 13; Length 393; Best Local Similarity 100.0%; Pred. No. 6, 7e-180; Matches 352; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AHQWVWPLNVRGRGDAILLTCAVPHPLIDITKULLAIFGPLMVQAGITKVPUFVR 60

Db 42 AHQWVWPLNVRGRGDAILLTCAVPHPLIDITKULLAIFGPLMVQAGITKVPUFVR 101

QY 61 AQGIRACMVRKAGGHYVQAFMKLAALTGTGVYDHUTPLQDWAHGLRDLAVAEPV 120

Db 102 AQGIRACMVRKAGGHYVQAFMKLAALTGTGVYDHUTPLQDWAHGLRDLAVAEPV 161

QY 121 IFSPDMEVKLTWTGADTAACGDIISGLPVAARRSHIELGPADNEFGQWRLLAPITAVSQ 180

Db 162 IFSPDMEVKLTWTGADTAACGDIISGLPVSARRGREGELGPADNEFGQWRLLAPITAVSQ 221

QY 181 QTRGLIGCITSLGRDKRQVEGVQVSTATOSFLATCNGVWTVHAGSKTLAQPK 240

Db 222 QTRGLIGCITSLGRDKRQVEGVQVSTATOSFLATCNGVWTVHAGSKTLAQPK 281

QY 241 GPTIQMYTNDQDQJGWQAPPGRSMPTCTCGSSDLYLVRHADWIPRRGDSRGSLIS 300

Db 282 GPTIQMYTNDQDQJGWQAPPGRSMPTCTCGSSDLYLVRHADWIPRRGDSRGSLIS 341

QY 301 PRPVSYLKSSGGPLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMETMR 352

Db 342 PRPVSYLKSSGGPLCPSGHAVGIFRAAVCTRGVAKAVDFIPVESMETMR 393

RESULT 7

US-10-017-736-2

; Sequence 2; Application US/10017736

; Publication No. US200201926401

; GENERAL INFORMATION:

; APPLICANT: Boehringer Ingelheim (Canada) Ltd.

; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease

; FILE REFERENCE: 13/082

; CURRENT APPLICATION NUMBER: US/10/017,736

; CURRENT FILING DATE: 2001-12-14

; PRIORITY NUMBER: 60/256,031

; NUMBER OF SEQ ID NOS: 21

; SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO 2

; LENGTH: 409

; TYPE: PRT

; ORGANISM: HCV

US-10-017-736-2

Query Match 100.0%; Score 1842; DB 13; Length 409; Best Local Similarity 100.0%; Pred. No. 7, 1e-180; Matches 352; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AHQWVWPLNVRGRGDAILLTCAVPHPLIDITKULLAIFGPLMVQAGITKVPUFVR 60

Db 46 AHQWVWPLNVRGRGDAILLTCAVPHPLIDITKULLAIFGPLMVQAGITKVPUFVR 105

QY 61 AQGIRACMVRKAGGHYVQAFMKLAALTGTGVYDHUTPLQDWAHGLRDLAVAEPV 120

RESULT 8

US-10-650-585-2

Sequence 2, Application US/10650585  
Publication No. US20040077066A1  
GENERAL INFORMATION:

APPLICANT: Boehringer Ingelheim (Canada) Ltd.  
TITLE OF INVENTION: Purified Active HCV NS2/3 Protease  
FILE REFERENCE: 13/082  
CURRENT APPLICATION NUMBER: US/10/017,736  
CURRENT FILING DATE: 2001-12-1  
PRIORITY APPLICATION NUMBER: 60/256,031  
PRIOR FILING DATE: 2000-12-15  
NUMBER OF SEQ ID NOS: 21  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 14  
LENGTH: 341  
TYPE: PRT  
ORGANISM: HCV  
US-10-017-736-14

Query Match

Best Local Similarity 96.5%; Score 1778; DB 13; Length 341;  
Matches 341; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 VRCGRDAAITLTCVAVHPELIFPDITKLLAIFGPLMVLQAGIKVYFRAQGILTRACMV  
Db 1 VRGGDRDAITLTCVAVHPELIFPDITKLLAIFGPLMVLQAGIKVYFRAQGILTRACMV  
QY 72 RKAAGGTVQMAFMKLAALTGTGVYDHLPQWAHGLRDLAVAVEFVIFSDMEVKLT 131  
Db 61 RKAAGGTVQMAFMKLAALTGTGVYDHLPQWAHGLRDLAVAVEFVIFSDMEVKLT 120  
QY 132 WGAATTAACGDIISGLPV/SARRGEITLGPADNFGQGRILLAPITAYSQQTSLIGCIT 191  
Db 121 WGAATTAACGDIISGLPV/SARRGRBILGPADNFEGQWRLLAPITAYSQQTSLIGCIT 180  
QY 192 SITGDRKNQVEGEVQVSTATOSFLATCUNGCVWTFHGAGSKTLAGKPTOMYTND 251  
Db 181 SITGDRKNQVEGEVQVSTATOSFLATCUNGCVWTFHGAGSKTLAGKPTOMYTND 240  
QY 252 QDLVGWQAPPGRASMTPCTCGSSDLYLVRHADVIPVRRGDSRGSLSPRPSYLKSS 311  
Db 241 QDLVGWQAPPGRASMTPCTCGSSDLYLVRHADVIPVRRGDSRGSLSPRPSYLKSS 300  
QY 312 GGPLICPSPSCHAVGIFRAAVCTRGVAKAVDFIPVESMETMR 352  
Db 301 GGPLICPSPSCHAVGIFRAAVCTRGVAKAVDFIPVESMETMR 341

RESULT 9

US-10-017-736-14

Sequence 14, Application US/10017736  
Publication No. US2001017736A1  
GENERAL INFORMATION:

APPLICANT: Boehringer Ingelheim (Canada) Ltd.  
TITLE OF INVENTION: Purified Active HCV NS2/3 Protease  
FILE REFERENCE: 13/082  
CURRENT APPLICATION NUMBER: US/10/017,736  
CURRENT FILING DATE: 2001-12-1  
PRIORITY APPLICATION NUMBER: 60/256,031  
PRIOR FILING DATE: 2000-12-15  
PRIORITY APPLICATION NUMBER: 60/256,031  
PRIOR FILING DATE: 2000-12-15  
NUMBER OF SEQ ID NOS: 21  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 14  
LENGTH: 341  
TYPE: PRT  
ORGANISM: HCV  
US-10-017-736-14

Query Match

Best Local Similarity 96.5%; Score 1778; DB 16; Length 341;  
Matches 341; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 VRCGRDAAITLTCVAVHPELIFPDITKLLAIFGPLMVLQAGIKVYFRAQGILTRACMV  
Db 1 VRGGDRDAITLTCVAVHPELIFPDITKLLAIFGPLMVLQAGIKVYFRAQGILTRACMV  
QY 72 RKAAGGTVQMAFMKLAALTGTGVYDHLPQWAHGLRDLAVAVEFVIFSDMEVKLT 131  
Db 61 RKAAGGTVQMAFMKLAALTGTGVYDHLPQWAHGLRDLAVAVEFVIFSDMEVKLT 120  
QY 132 WGAATTAACGDIISGLPV/SARRGEITLGPADNFGQGRILLAPITAYSQQTSLIGCIT 191  
Db 121 WGAATTAACGDIISGLPV/SARRGRBILGPADNFEGQWRLLAPITAYSQQTSLIGCIT 180  
QY 192 SITGDRKNQVEGEVQVSTATOSFLATCUNGCVWTFHGAGSKTLAGKPTOMYTND 251  
Db 181 SITGDRKNQVEGEVQVSTATOSFLATCUNGCVWTFHGAGSKTLAGKPTOMYTND 240  
QY 252 QDLVGWQAPPGRASMTPCTCGSSDLYLVRHADVIPVRRGDSRGSLSPRPSYLKSS 311  
Db 241 QDLVGWQAPPGRASMTPCTCGSSDLYLVRHADVIPVRRGDSRGSLSPRPSYLKSS 300  
QY 312 GGPLICPSPSCHAVGIFRAAVCTRGVAKAVDFIPVESMETMR 352  
Db 301 GGPLICPSPSCHAVGIFRAAVCTRGVAKAVDFIPVESMETMR 341

RESULT 10

US-10-650-585-14

Sequence 14, Application US/10650585  
Publication No. US20040077066A1  
GENERAL INFORMATION:

APPLICANT: Boehringer Ingelheim (Canada) Ltd.  
TITLE OF INVENTION: Purified Active HCV NS2/3 Protease  
FILE REFERENCE: 13/082  
CURRENT APPLICATION NUMBER: US/10/017,736  
CURRENT FILING DATE: 2001-08-28  
PRIORITY APPLICATION NUMBER: US/10/017,736A  
PRIOR FILING DATE: 2001-12-14  
PRIORITY APPLICATION NUMBER: 60/256,031  
PRIOR FILING DATE: 2000-12-15  
NUMBER OF SEQ ID NOS: 21  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 14  
LENGTH: 341  
TYPE: PRT  
ORGANISM: HCV  
US-10-650-585-14

Query Match

Best Local Similarity 96.5%; Score 1778; DB 16; Length 341;  
Matches 341; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 VRCGRDAAITLTCVAVHPELIFPDITKLLAIFGPLMVLQAGIKVYFRAQGILTRACMV  
Db 1 VRGGDRDAITLTCVAVHPELIFPDITKLLAIFGPLMVLQAGIKVYFRAQGILTRACMV  
QY 72 RKAAGGTVQMAFMKLAALTGTGVYDHLPQWAHGLRDLAVAVEFVIFSDMEVKLT 131  
Db 61 RKAAGGTVQMAFMKLAALTGTGVYDHLPQWAHGLRDLAVAVEFVIFSDMEVKLT 120  
QY 132 WGAATTAACGDIISGLPV/SARRGEITLGPADNFGQGRILLAPITAYSQQTSLIGCIT 191  
Db 121 WGAATTAACGDIISGLPV/SARRGRBILGPADNFEGQWRLLAPITAYSQQTSLIGCIT 180  
QY 192 SITGDRKNQVEGEVQVSTATOSFLATCUNGCVWTFHGAGSKTLAGKPTOMYTND 251  
Db 181 SITGDRKNQVEGEVQVSTATOSFLATCUNGCVWTFHGAGSKTLAGKPTOMYTND 240  
QY 252 QDLVGWQAPPGRASMTPCTCGSSDLYLVRHADVIPVRRGDSRGSLSPRPSYLKSS 311  
Db 241 QDLVGWQAPPGRASMTPCTCGSSDLYLVRHADVIPVRRGDSRGSLSPRPSYLKSS 300  
QY 312 GGPLICPSPSCHAVGIFRAAVCTRGVAKAVDFIPVESMETMR 352  
Db 301 GGPLICPSPSCHAVGIFRAAVCTRGVAKAVDFIPVESMETMR 345



OTHER INFORMATION: :

US-09-919-901-4

Query Match 95.4%; Score 1757; DB 10; length 1692;  
Best Local Similarity 93.8%; Pred. No. 2.7e-170;  
Matches 330; Conservative 12; Mismatches 10; Indels 0; Gaps 0;

APPLICANT: De Francesco, Raffaele  
Applicant: Migliaccio, Giovanni  
TITLE OF INVENTION: HEPATITIS C VIRUS REPLICONS AND REPLICON  
TITLE OF INVENTION: ENHANCED CELLS  
FILE REFERENCE: ITR003P

CURRENT FILING DATE: 2003-07-21  
PRIOR APPLICATION NUMBER: PCT/EP02/00526  
PRIOR FILING DATE: 2002-01-16  
PRIOR APPLICATION NUMBER: 60/263,479  
PRIOR FILING DATE: 2001-01-23  
NUMBER OF SEQ ID NOS: 13

SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 1  
LENGTH: 3010  
TYPE: RT  
ORGANISM: Con 1 HCV isolate nucleic acid  
US-10-467-000-1

Query Match 95.9%; Score 1766; DB 12; length 3010;  
Best Local Similarity 94.3%; Pred. No. 7.4e-171;  
Matches 332; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

QY 1 AHLQWIPPLNVRGRDATTILTCAVHPELIFDTKLIAIFGPMVHQAGITKVPYFV 60  
Db 1 AHLQWIPPLNVRGRDATTILTCAVHPELIFDTKLIAIFGPMVHQAGITKVPYFV 60  
QY 61 AQLIRACMVRKAAGGYQMAFMKLAIGTGYWDHTPLQDWAHGLDLAVAEPV 120  
Db 194 AQLIRACMVRKAAGGYQMAFMKLAIGTGYWDHTPLQDWAHGLDLAVAEPV 120  
Db 121 IFSDMEVKITWGADTAACCDISGLPVSARGREILLGPADNFEGQWRLLAPIASQ 180  
Db 254 VFSDEMETKLTITWGADTAACCDISGLPVSARGREILLGPADSLERGWRLLAPIASQ 313  
QY 181 QTRGLIGCITSLSLGRDKRDNQVEGEVQVNVSTATOSFLATCVNGCWTWFGASSTKLAGPK 240  
Db 314 QTRGLIGCITSLSLGRDKRDNQVEGEVQVNVSTATOSFLATCVNGCWTWFGASSTKLAGPK 240  
QY 241 GPITQMTNTQDLYGWAQPGASMTPTCGSDDLYLVTRADVIPRERRGSRSGSLS 300  
Db 374 GPITQMTNTQDLYGWAQPGASMTPTCGSDDLYLVTRADVIPRERRGSRSGSLS 433  
QY 301 PRPVSYLKGSGGPLCPSPCHAVGIFRAVCTRGAVAKVDPIVESMETMR 352  
Db 434 PRPVSYLKGSGGPLCPSPCHAVGIFRAVCTRGAVAKVDPIVESMETMR 485

QY 121 IFSDMEVKITWGADTAACCDISGLPVSARGREILLGPADNFEGQWRLLAPIASQ 180  
Db 975 VFSDEMETKLTITWGADTAACCDISGLPVSARGREILLGPADSLERGWRLLAPIASQ 1034  
QY 181 QTRGLIGCITSLSLGRDKRDNQVEGEVQVNVSTATOSFLATCVNGCWTWFGASSTKLAGPK 240  
Db 1095 GPITQMTNTQDLYGWAQPGASMTPTCGSDDLYLVTRADVIPRERRGSRSGSLS 1154  
QY 301 PRPVSYLKGSGGPLCPSPCHAVGIFRAVCTRGAVAKVDPIVESMETMR 352  
Db 1155 PRPVSYLKGSGGPLCPSPCHAVGIFRAVCTRGAVAKVDPIVESMETMR 1206

RESULT 14

US-10-919-901-4

Sequence 4, Application US/10191966

Publication No. US20030175692A1

GENERAL INFORMATION:

APPLICANT: Potts, Karen E.

APPLICANT: Jackson, Roberta L.

APPLICANT: Patrick, Amy K.

TITLE OF INVENTION: REPORTER GRNE SYSTEM FOR USE IN CELL-BASED ASSESSMENT

TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE

FILE REFERENCE: 0125-005A

CURRENT APPLICATION NUMBER: US10/191, 966

CURRENT FILING DATE: 2002-07-10

PRIOR APPLICATION NUMBER: US09/263, 933

PRIOR FILING DATE: 1999-01-08

PRIOR APPLICATION NUMBER: 09/129, 611

PRIOR FILING DATE: 1998-08-05

NUMBER OF SEQ ID NOS: 33

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 4  
LENGTH: 1692  
TYPE: RT  
ORGANISM: Artificial Sequence

FEATURE: OTHER INFORMATION: :

US-10-191-966-4

Query Match 95.4%; Score 1757; DB 14; length 1692;  
Best Local Similarity 93.8%; Pred. No. 2.7e-170;  
Matches 330; Conservative 12; Mismatches 10; Indels 0; Gaps 0;

QY 1 AHLQWIPPLNVRGRDATTILTCAVHPELIFDTKLIAIFGPMVHQAGITKVPYFV 60  
Db 134 AHLQWIPPLNVRGRDATTILTCAVHPELIFDTKLIAIFGPMVHQAGITKVPYFV 193  
QY 61 AQLIRACMVRKAAGGYQMAFMKLAIGTGYWDHTPLQDWAHGLDLAVAEPV 120  
Db 194 AQLIRACMVRKAAGGYQMAFMKLAIGTGYWDHTPLQDWAHGLDLAVAEPV 253  
QY 121 IFSDMEVKITWGADTAACCDISGLPVSARGREILLGPADNFEGQWRLLAPIASQ 180

Fri May 7 13:37:03 2004

us-10-650-585-13.rapb

Page 7

Db 254 VFSDMETKIIIWGADTAACDDIILGLPVSAARGKEILIGPADSIEGRGRWRLIAPITASQ 313  
Qy 181 CTRGLIGCITISLICRDKNQVEGEVQVSTATOSFLACYNGCYCWTYHGAASKTLAGPK 240  
Db 314 CTRGLIGCITISLICRDKNQVEGEVQVSTATOSFLATCVNGYCWTYHGAASKTLAGPK 373  
Qy 241 GRITOMYTNVDQDLYVGWQAPPGRASMTPTCCGSSDIYLUTRAHADVPRRRDSRGSLIS 300  
Db 374 GRITOMYTNVDQDLYVGWQAPPGRASLTCTCGSSDLYLUTRAHADVPRRRDSRGSLIS 433  
Qy 301 PRPVSYLKGSGGPLLCPSGHAVGIFRAAVCTRGYAKAVDFIPVESMETMR 352  
Db 434 PRPVSYLKGSGGPLLCPSGHAVGIFRAAVCTRGYAKAVDFPVVESMETMR 485

Search completed: May 6, 2004, 09:43:21  
Job time : 37.0084 secs



GenCore version 5.1.6  
 Copyright (c) 1993 - 2004 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: May 6, 2004, 09:08:45 ; Search time 46,4159 Seconds  
 (without alignments)  
 2075.771 Million cell updates/sec

Title: US-10-650-585-14  
 Perfect score: 1778  
 Sequence: VRGGRGAAAILITCAVHPELI.....RGVAVADFLPVESMETMVR 341

Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext: 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries

Database : A\_Geneseq\_29Jan04; \*

- 1: geneseq1900; \*
- 2: geneseq1990s; \*
- 3: geneseq2000s; \*
- 4: geneseq2001s; \*
- 5: geneseq2002s; \*
- 6: geneseq2003s; \*
- 7: geneseq2004s; \*
- 8: geneseq2004s; \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

## %

| Result No. | Score | Query Match Length | DB ID | Description |
|------------|-------|--------------------|-------|-------------|
| 1          | 1778  | 100.0              | 341   | ABG32187    |
| 2          | 1778  | 100.0              | 352   | ABG32186    |
| 3          | 1778  | 100.0              | 380   | ABG32185    |
| 4          | 1778  | 100.0              | 393   | ABG32184    |
| 5          | 1778  | 100.0              | 409   | ABG32181    |
| 6          | 1710  | 96.9               | 3010  | AAR68624    |
| 7          | 1705  | 95.9               | 3010  | AAR68622    |
| 8          | 1704  | 95.8               | 3010  | AAR68664    |
| 9          | 1702  | 95.7               | 201   | ABG30501    |
| 10         | 1702  | 95.7               | 201   | ABG30591    |
| 11         | 1702  | 95.7               | 2201  | ABG30600    |
| 12         | 1702  | 95.7               | 2201  | ABG30581    |
| 13         | 1702  | 95.7               | 2201  | ABG30593    |
| 14         | 1702  | 95.7               | 2201  | ABG30582    |
| 15         | 1702  | 95.7               | 2201  | ABG30580    |
| 16         | 1702  | 95.7               | 201   | ABG30587    |
| 17         | 1702  | 95.7               | 2201  | ABG30599    |
| 18         | 1702  | 95.7               | 2201  | ABG30594    |
| 19         | 1702  | 95.7               | 2201  | ABG30598    |
| 20         | 1702  | 95.7               | 2201  | ABG30595    |
| 21         | 1702  | 95.7               | 3010  | ABG32458    |
| 22         | 1702  | 95.7               | 3010  | ABG32459    |
| 23         | 1702  | 95.7               | 3010  | ABG32451    |
| 24         | 1702  | 95.7               | 3010  | ABG32455    |
| 25         | 1702  | 95.7               | 3010  | ABG32457    |

## ALIGNMENTS

|          |             |  |
|----------|-------------|--|
| RESULT 1 | ABG32187    | ID ABG32187 standard; protein; 341 AA.   |
| XX       | ABG32187;   | XX   |
| XX       | 05-NOV-2002 | (first entry)  |
| XX       | DE          | HCV protease NS2/3 truncation mutant 866-1205.   |
| XX       | XX          | XX; enzyme; protease; NS2/3 (866-1206); hepatitis C virus infection; chronic liver disease; cirrhosis; end-stage liver disease; virucide; hepatotropic; antiinflammatory; lauryldiethyamine oxide; LDRO; chaotropic agent; mutant; mutein.   |
| XX       | XX          | XX; Hepatitis C virus.   |
| OS       | OS          | Synthetic.   |
| XX       | PN          | WO200248375-A2.  |
| XX       | PD          | 20-JUN-2002.   |
| XX       | PF          | 13-DEC-2001; 2001WO-CA001796.  |
| XX       | PR          | 15-DEC-2000; 2000US-025631P.   |
| XX       | PA          | (BOEH ) BOHRINGER INGELHEIM CANADA LTD.  |
| XX       | PT          | Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;  |
| XX       | PI          | DR WBI; 2002-599511/64.  |
| XX       | PS          | Claim 41; Page 62-63; 67EP; English.   |
| XX       | PT          | Novel polypeptide for screening inhibitors of non-structural proteases useful as therapeutic agents against hepatitis C virus, comprises full length non-structural protease, or its truncation.   |
| XX       | PS          | The invention relates to an isolated polypeptide consisting of a full-length HCV (hepatitis C virus) non-structural (NS) 2/3 protease (referred to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal residue amino acid 810 to 906, or having a minimal amino acid sequence from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length NS2/3 protease. Also included are (1) a composition (C) comprising an isolated HCV NS2/3 protease selected from full length NS2/3 protease, or its truncation, or a mutated sequence, where the protease is in a solution comprising a sufficient concentration of lauryldiethyamine oxide (LDAO) to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide |

CC appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3 protease, involving isolating the protease in the presence of a chaotropic agent, refolding the isolated protease by contacting it with a reducing agent, and LDAO in the presence of reduced concentration of the chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3 protease, involving diluting refolded, inactive NS2/3 protease in a medium containing an activation detergent to induce auto-cleavage of the NS2/3 protease; (5) measuring (M3) the auto-cleavage activity of NS2/3 protease, involving incubating the active NS2/3 protease produced by M2 for sufficient time to induce auto-cleavage of NS2/3 protease and produce cleavage products or their fragments, and measuring the presence or absence of uncleaved NS2/3 protease, cleavage products or their fragments ; and (6) screening a potential inhibitor of auto-cleavage activity of an active NS2/3 protease, involving carrying out M3 in the presence of, or absence of the potential inhibitor, comparing the amount of uncleaved NS2/3 protease, cleavage products or their fragments, and measuring the presence or absence of uncleaved NS2/3 protease, cleavage products or their fragments useful for detailed biochemical characterisation of the enzymes and in the development of *in vitro* assays for screening novel inhibitors of NS2/3 protease which are useful as therapeutic agents against HCV infection (which causes chronic liver disease, cirrhosis and end-stage liver disease. M1 is useful for high level production of protease. The present sequence represents the NS2/3 truncation mutant 866-1206 (numbered relative to the full length NS2/3 protein)

XX Sequence 341 AA;

Query Match Score 1778; DB 5; Length 341;  
Best Local Similarity 100.0%; Pred. No. 9\_5e-164; Mismatches 0; Indels 0; Gaps 0;  
Matches 341; Conservative 0; MisMatches 0; Del 0; Insert 0;

Qy 1 VRGGDAAILLTCAVPHELFDITKLLAIFGPMVLQAGITKPYFVRAGLIRACMLV 60  
Db 1 VRGGDAAILLTCAVPHELFDITKLLAIFGPMVLQAGITKPYFVRAGLIRACMLV 60

Qy 61 RKAAGHYYVQAFMKLAAIGTGYVVDLTLQDWAHGIRDLAVALVEPVIFSDMEVKIT 120  
Db 61 RKAAGHYYVQAFMKLAAIGTGYVVDLTLQDWAHGIRDLAVALVEPVIFSDMEVKIT 120

Qy 121 WGADTAAGDTISGIPVSARSGRETLGLGADNFFGQGRLLAPITASQQTREGLGICIT 180  
Db 121 WGADTAAGDTISGIPVSARSGRETLGLGADNFFGQGRLLAPITASQQTREGLGICIT 180

Qy 181 SLTGDRKNQVEGEVQVSTATOSFLATCNGVCMWVFGAGSKTLAGPKGPITQMYNTVD 240  
Db 181 SLTGDRKNQVEGEVQVSTATOSFLATCNGVCMWVFGAGSKTLAGPKGPITQMYNTVD 240

Qy 241 QDLVQWQAPRGARSMPCTCGSSDLYLTVRHADTPVRRGDSRSGLSPRPSYLGSS 300  
Db 241 QDLVQWQAPRGARSMPCTCGSSDLYLTVRHADTPVRRGDSRSGLSPRPSYLGSS 300

Qy 301 GGPILCPSGHANGVFRAAVCTRGYAKAVDPRIPESMETMR 341  
Db 301 GGPILCPSGHANGVFRAAVCTRGYAKAVDPRIPESMETMR 341

RESULT 2  
ABG32186 ID ABG32186 standard; protein; 352 AA.  
XX AC ABG32186;  
XX DT 05-NOV-2002 (first entry)  
XX HCV protease NS2/3 truncation mutant 855-1206.  
XX HCV; enzyme; protease; NS2/3 (855-1206); hepatitis C virus infection; chronic liver disease; cirrhosis; end-stage liver disease; virucide; hepatotrophic; antiinflammatory; lauryldiethylamine oxide; LDAO; chaotropic agent; mutant; muttein.  
XX OS Hepatitis C virus.  
OS Synthetic.  
XX

PN WO200248375-A2.  
XX 20-JUN-2002.  
PF XX 13-DEC-2001; 2001WO-CA001796.  
PR XX 15-DEC-2000; 2000US-0256031P.  
PA (BOEHRINGER INGELHEIM CANADA LTD.  
XX PI Thibeault D, Iammarre D, Maurice R, Pilote L, Pause A;  
XX DR XX WPI; 2002-599511/64.  
XX  
PT Novel polypeptide for screening inhibitors of non-structural proteases useful as therapeutic agents against hepatitis C virus, comprises full length non-structural protease, or its truncation.  
CC Claim 41; Page 61-62; 67PP; English.  
CC The invention relates to an isolated polypeptide consisting of a full-length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal residue amino acid 810 to 905, or having a minimal amino acid sequence from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length NS2/3 protease. Also included are (1) a composition (C) comprising an isolated HCV NS2/3 protease selected from full length NS2/3 protease, or its truncation or a mutated sequence, where the protease is in a solution comprising a sufficient concentration of lauryldiethylamine oxide (LDAO) to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3 protease, involving isolating the protease in the presence of a chaotropic agent, refolding the isolated protease by contacting it with a reducing agent, and LDAO in the presence of reduced concentration of the chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3 protease, involving diluting refolded, inactive NS2/3 protease in a medium containing an activation detergent to induce auto-cleavage activity of NS2/3 protease; (5) measuring (M3) the auto-cleavage activity of NS2/3 protease, involving incubating the active NS2/3 protease produced by M2 for sufficient time to induce auto-cleavage of NS2/3 protease and produce cleavage products or their fragments, and measuring the presence or absence of uncleaved NS2/3 protease, cleavage products or their fragments ; and (6) screening a potential inhibitor of auto-cleavage activity of an active NS2/3 protease, involving carrying out M3 in the presence of, or absence of the potential inhibitor, comparing the amount of uncleaved NS2/3 protease, cleavage products or their fragments. The protease is useful for detailed biochemical characterisation of the enzymes and in the development of *in vitro* assays for screening novel inhibitors of NS2/3 protease which are useful as therapeutic agents against HCV infection (which causes chronic liver disease, cirrhosis and end-stage liver disease. M1 is useful for high level production of protease. The present sequence represents the NS2/3 truncation mutant 855-1206 (numbered relative to the full length NS2/3 protein)

XX Sequence 352 AA;

Query Match Score 1778; DB 5; Length 352;  
Best Local Similarity 100.0%; Pred. No. 1e-163; Mismatches 0; Indels 0; Gaps 0;  
Matches 341; Conservative 0; MisMatches 0; Del 0; Insert 0;

Qy 1 VRGGDAAILLTCAVPHELFDITKLLAIFGPMVLQAGITKPYFVRAGLIRACMLV 60  
Db 1 VRGGDAAILLTCAVPHELFDITKLLAIFGPMVLQAGITKPYFVRAGLIRACMLV 71

Qy 61 RKAAGHYYVQAFMKLAAIGTGYVVDLTLQDWAHGIRDLAVALVEPVIFSDMEVKIT 120  
Db 72 RKAAGHYYVQAFMKLAAIGTGYVVDLTLQDWAHGIRDLAVALVEPVIFSDMEVKIT 131

Qy 121 WGADTAAGDTISGIPVSARSGRETLGLGADNFFGQGRLLAPITASQQTREGLGICIT 180  
Db 132 WGADTAAGDTISGIPVSARSGRETLGLGADNFFGQGRLLAPITASQQTREGLGICIT 191

Qy 181 SLTGDRKNQVEGEVQVSTATOSFLATCNGVCMWVFGAGSKTLAGPKGPITQMYNTVD 240

|    |     |   |     |
|----|-----|---|-----|
| Db | 192 | SLTGRDKNQVEGEVQVSTATQSQFLATCNGVCWTVFHGAGSKTLAQPKGPITQMTND   | 251 |
| QY | 241 | QDLVGWQAPPGCARSMTPCTCGSSDLYLIVTRHADTVFVRRGDSRSGLSPRVSYIKGSS | 300 |
| Db | 252 | QDLVGWQAPPGARSMTPCTCGSSDLYLIVTRHADTVFVRRGDSRSGLSPRVSYIKGSS  | 311 |
| QY | 301 | GGPLLCPSGHAVGFRAVCTRGAVAKADVDFPVESEMMR 341                  |     |
| Db | 312 | GGPLLCPSGHAVGIRRAVCTRGAVAKADVDFPVESEMMR 352                 |     |

active NS2/3 protease, involving carrying out *in vitro* assays in the presence or absence of the potential inhibitor, comparing the amount of uncleaved NS2/3 protease, cleavage products or their fragments. The protease is useful for detailed biochemical characterization of the enzymes and in the development of *in vitro* assays for screening novel inhibitors of NS2/3 protease which are useful as therapeutic agents against HCV infection (which causes chronic liver disease, cirrhosis and end-stage liver disease). M1 is useful for high level production of protease. The present sequence represents the NS2/3 truncation mutant 827-1206 (numbered relative to the full length NS2/3 protein).

|        |  |
|--------|--|
| RESULT | 3  |
| ID     | ABG32185 standard; protein; 380 AA.  |
| XX     |  |
| AC     | ABC32185;  |
| DT     | 05-NOV-2002 (first entry)  |
| XX     |  |
| DE     | HCV protease NS2/3 truncation mutant   |
| XX     |  |
| KW     | HCV; enzyme; protease; NS2/3 (827-1111); chronic liver disease; cirrhosis; end-stage hepatotropic; antiinflammatory; lauroyl chacterropic agent; mutant; mutein. |
| KW     |  |
| OS     | Repetitis C virus.   |
| OS     | Synthetic.   |
| XX     |  |
| PN     | WO200249375-A2.  |
| XX     |  |
| PD     | 20-JUN-2002.   |
| XX     |  |
| PF     | 13-DEC-2001; 2001WO-CA011796.  |
| XX     |  |
| PR     | 15-DEC-2000; 2000US025631P.  |
| XX     |  |
| PA     | (BOHR ) BOEHRINGER INGELHEIM CANADA  |
| XX     |  |
| PI     | Thibeault D, Lamarre D, Maurice R  |
| DR     | WPI, 2002-599511/64.   |
| XX     |  |
| PT     | Novel polypeptide for screening inhibitors useful as therapeutic agents against hepatitis C virus.   |
| PT     | length non-structural protease, or   |
| XX     |  |
| PS     | Claim 41; Page 60-61; 672P; English  |

| Query Match           |             | 100.0%  | Score 1778;  | DB 5;     | Length 380;       |
|-----------------------|-------------|---|--------------|-----------|-------------------|
| Best Local Similarity |             | 100.0%;   | Pred. No.    | 1.1e-163; |                   |
| Matches               |             | 341;  | Conservative | 0;        | Mismatches        |
|                       |             |   |              |           | Indels 0; Gaps 0; |
| OY                    | 1           | VRCGGRDAILLTCAVHPPELFDITKLLIAIFGLRMLMVIQAGITKVYFVRAGQLIRACMLV | 60           |           |                   |
| Db                    | 40          | VRCGGRDAILLTCAVHPPELFDITKLLIAIFGLRMLMVIQAGITKVYFVRAGQLIRACMLV | 99           |           |                   |
| QY                    | 61          | RKAAGHTYVOMAMKMLAALTGTYYDHLTPDWAHGLDAAVAVFVFSDEVKIT           | 120          |           |                   |
| Db                    | 100         | RKAAGHTYVOMAMKMLAALTGTYYDHLTPDWAHGLDAAVAVFVFSDEVKIT           | 159          |           |                   |
| QY                    | 121         | WGADTAAGDITSGIPIVSARRGEELIGPADNFEQGQRLLAPITAVSQRTGRLGCIT      | 180          |           |                   |
| Db                    | 160         | WGADTAACGDIISGIPVSARRGRBLLIGPADNFEQGQRLLAPITAVSQRTGRLGCIT     | 219          |           |                   |
| QY                    | 181         | SLGGRDKNOVEGVQVNSTATOSPLATCNGVWTFHGASKTLAPKGPTIQMYNTD         | 240          |           |                   |
| Db                    | 220         | SLGGRDKNOVEGEVQVNSTATOSPLATCNGVWTFHGASKTLAPKGPTIQMYNTD        | 279          |           |                   |
| QY                    | 241         | QDLVGWQAPPGAASMTPCTCGSSDLILVTRHADIVPVRRGDSRSGLSPRVSYLKSS      | 300          |           |                   |
| Db                    | 280         | QDLVGWQAPPGARSMTPCTCGSSDLILVTRHADIVPVRRGDSRSGLSPRVSYLKSS      | 339          |           |                   |
| QY                    | 301         | GGFLLCPSGHAVGFRAVCVTRSYKAKDFIPFESMETMR                        | 341          |           |                   |
| Db                    | 340         | GGFLLCPSGHAVGFRAVCVTRSYKAKDFIPFESMETMR                        | 380          |           |                   |
| RESULT 4              |             |   |              |           |                   |
| ID                    | ABG32184    | ABG32184 standard; protein; 393 AA.                           |              |           |                   |
| XX                    |             |   |              |           |                   |
| AC                    | ABG32184;   |   |              |           |                   |
| XX                    |             |   |              |           |                   |
| DT                    | 05-NOV-2002 | (first entry)   |              |           |                   |
| XX                    |             |   |              |           |                   |

XX  
PT Novel polypeptide for screening inhibitors of non-structural proteases  
PT useful as therapeutic agents against hepatitis C virus, comprises full  
length non-structural protease, or its truncation.  
XX  
PS Claim 41; Page 59 60; 67pp; English.

The invention relates to an isolated polypeptide consisting of a full-length HCV (hepatitis C virus) non-structural (NS) 2/3 protease (referred to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal residue amino acid 810 to 906, or having a minimal amino acid sequence from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length NS2/3 protease. Also included are (1) a composition (C) comprising an isolated HCV NS2/3 protease selected from full length NS2/3 protease, or its truncation or a mutated sequence where the protease is in a solution comprising a sufficient concentration of lauryldiethylamine oxide (LDAO) to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide appearing as ABG2198; (3) producing (M1) a refolded, inactive HCV NS2/3 protease, involving isolating the protease in the presence of a chaotropic agent, refolding the isolated protease by contacting it with a reducing agent, and LDAO in the presence of reduced concentration of the chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3 protease, involving diluting refolded inactive NS2/3 protease in a medium containing an activation detergent to induce auto-cleavage of the NS2/3 protease; (5) measuring (M3) the auto-cleavage activity of NS2/3 protease, involving incubating the active NS2/3 protease produced by M2 for sufficient time to induce auto-cleavage of NS2/3 protease and produce cleavage products or their fragments, and measuring the presence or absence of uncleaved NS2/3 protease, cleavage products or their fragments ; and (6), screening a potential inhibitor of auto-cleavage activity of an active NS2/3 protease, involving carrying out M3 in the presence of, or absence of the potential inhibitor, comparing the amount of uncleaved or NS2/3 protease, cleavage products or their fragments. The protease is useful for detailed biochemical characterisation of the enzymes and in the development of *in vitro* assays for screening novel inhibitors of NS2/3 protease which are useful as therapeutic agents against HCV infection (which causes chronic liver disease, cirrhosis and end-stage liver disease). M1 is useful for high level production of protease. The present sequence represents the NS2/3 truncation mutant 815-1206 (numbered relative to the full length NS2/3 protein)

XX  
SQ Sequence 393 AA;

| Query Match           | Score  | DB | Length | Score  | DB | Length | Score      | DB | Length |
|-----------------------|--------|----|--------|--------|----|--------|------------|----|--------|
| Best Local Similarity | 100.0% | 5  | 393    | 1778   | 5  | 393    | 100.0%     | 0  | 0      |
| Matches               |        |    |        | Pred.  |    |        | Mismatches |    |        |
| 341;                  |        |    |        | No.    |    |        | 0;         |    |        |
| Conservative          |        |    |        | Indels |    |        | Caps       |    |        |
|                       |        |    |        | 0;     |    |        | 0;         |    |        |

Qy 1 VRGERDAILLTCAVHPPELFDTKLIAIFGPMVQAGITKPVYFVRAGILRAGLY 60  
Db 53 VRRGRDAILLTCAVHPPELFDTKLIAIFGPMVQAGITKPVYFVRAGILRAGLY 112  
Qy 61 RKAAGHGYQMAFMKGLAIGTGYVYDHUTPLQWAHAGIRDLAVLAVERVFSMEVKIT 120  
Db 113 RKAAGHGYQMAFMKGLAIGTGYVYDHUTPLQWAHAGIRDLAVLAVERVFSMEVKIT 172  
Qy 121 WGADTAACGDIISGIPVSARGRERLILGADNTEGOGWRLAPITASQQQRGLLGIT 180  
Db 173 WGADTAACGDIISGIPVSARGRERLILGADNTEGOGWRLAPITASQQQRGLLGIT 232  
Qy 181 SLTGDKDKNQEGEYQVWSTATOSFLATCNGVCMVFGAGSKILLAGRKRTOMYND 240  
Db 233 SLTGDKDKNQEGEYQVWSTATOSFLATCNGVCMVFGAGSKILLAGRKRTOMYND 292  
Qy 241 QDLYGWOARPGAREMTPTCGSSILYLVRHADYVPIVERRGDRGSLSPREVSYLKGS 300  
Db 293 QDLVGWQAFPGRASMTPTCGSSILYLVRHADYVPIVERRGDRGSLSPREVSYLKGS 352  
301 GGPILCPSPGHAVGIFRAAVCTRGVAKAUDFIPVESMETMR 341  
Qy 353 GGPILCPSPGHAVGIFRAAVCTRGVAKAUDFIPVESMETMR 393

XX  
PT Novel polypeptide for screening inhibitors of non-structural proteases  
PT useful as therapeutic agents against hepatitis C virus, comprises full  
length non-structural protease, or its truncation.  
XX  
PS Claim 41; Page 59 60; 67pp; English.

The invention relates to an isolated polypeptide consisting of a full-length HCV (hepatitis C virus) non-structural (NS) 2/3 protease (referred to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal residue amino acid 810 to 906, or having a minimal amino acid sequence from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length NS2/3 protease. Also included are (1) a composition (C) comprising an isolated HCV NS2/3 protease selected from full length NS2/3 protease, or its truncation or a mutated sequence, where the protease is in a solution comprising a sufficient concentration of lauryldiethylamine oxide (LDAO) to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide appearing as ABG2198; (3) producing (M1) a refolded, inactive HCV NS2/3 protease, involving isolating the protease in the presence of a chaotropic agent, refolding the isolated protease by contacting it with a reducing agent, and LDAO in the presence of reduced concentration of the chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3 protease, involving diluting refolded inactive NS2/3 protease in a medium containing an activation detergent to induce auto-cleavage of the NS2/3 protease; (5) measuring (M3) the auto-cleavage activity of NS2/3 protease, involving incubating the active NS2/3 protease produced by M2 for sufficient time to induce auto-cleavage of NS2/3 protease and produce cleavage products or their fragments, and measuring the presence or absence of uncleaved NS2/3 protease, cleavage products or their fragments ; and (6), screening a potential inhibitor of auto-cleavage activity of an active NS2/3 protease, involving carrying out M3 in the presence of, or absence of the potential inhibitor, comparing the amount of uncleaved NS2/3 protease, cleavage products or their fragments. The protease is useful for detailed biochemical characterisation of the enzymes and in the development of *in vitro* assays for screening novel inhibitors of NS2/3 protease which are useful as therapeutic agents against HCV infection (which causes chronic liver disease, cirrhosis and end-stage liver disease). M1 is useful for high level production of protease. The

ARG32181  
ID ABG32181 standard; protein; 409 AA.  
XX  
AC ABG32181;  
XX DT 05-NOW-2002 (first entry)  
XX DE HCV Protease NS2/3 (810-1206).  
XX KW HCV; enzme; protease; NS2/3 (810-1206); hepatitis C virus infection; chronic liver disease; cirrhosis; end-stage liver disease; viricide; hepatic; antiinflammatory; lauryldiethylamine oxide; LDAO; chaotropic agent; mutant; mutein.  
XX OS Hepatitis C virus.  
XX Synthetic.

XX Key Peptide Location/Qualifiers  
XX 398 - 409 /note= "Streptavidin tag"  
XX PR 13-DCC-2001; 2001WO-CA001796.  
XX PR 15-DCC-2000; 2000US-0256031P.  
XX DN WO200248375-A2.  
XX PA (BOEH ) BOHRINGER INGELHEIM CANADA LTD.  
XX PI Thibaault D, Lamarte D, Maurice R, Pilote L, Pause A;  
XX DR WPI; 2002-599511/64.  
XX DR N-PSPB; ABKG006.

XX PT Novel polypeptide for screening inhibitors of non-structural proteases  
PT useful as therapeutic agents against hepatitis C virus, comprises full  
length non-structural protease, or its truncation.  
XX  
PS Claim 42; Fig 1B; 67pp; English.

The invention relates to an isolated polypeptide consisting of a full-length HCV (hepatitis C virus) non-structural (NS) 2/3 protease (referred to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal residue amino acid 810 to 906, or having a minimal amino acid sequence from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length NS2/3 protease. Also included are (1) a composition (C) comprising an isolated HCV NS2/3 protease selected from full length NS2/3 protease, or its truncation or a mutated sequence, where the protease is in a solution comprising a sufficient concentration of lauryldiethylamine oxide (LDAO) to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide appearing as ABG2198; (3) producing (M1) a refolded, inactive HCV NS2/3 protease, involving isolating the protease in the presence of a chaotropic agent, refolding the isolated protease by contacting it with a reducing agent, and LDAO in the presence of reduced concentration of the chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3 protease, involving diluting refolded inactive NS2/3 protease in a medium containing an activation detergent to induce auto-cleavage of the NS2/3 protease; (5) measuring (M3) the auto-cleavage activity of NS2/3 protease, involving incubating the active NS2/3 protease produced by M2 for sufficient time to induce auto-cleavage of NS2/3 protease and produce cleavage products or their fragments, and measuring the presence or absence of uncleaved NS2/3 protease, cleavage products or their fragments ; and (6), screening a potential inhibitor of auto-cleavage activity of an active NS2/3 protease, involving carrying out M3 in the presence of, or absence of the potential inhibitor, comparing the amount of uncleaved NS2/3 protease, cleavage products or their fragments. The protease is useful for detailed biochemical characterisation of the enzymes and in the development of *in vitro* assays for screening novel inhibitors of NS2/3 protease which are useful as therapeutic agents against HCV infection (which causes chronic liver disease, cirrhosis and end-stage liver disease). M1 is useful for high level production of protease. The

CC present sequence represents the NS2/3 (810-1206) protein, which has a C-terminal streptavidin tag  
 XX SQ Sequence 409 AA;

Query Match 100.0%; Score 1778; DB 5; Length 409;

Best Local Similarity 100.0%; Pred. No. 1.2e-163; Mismatches 0; Indels 0; Gaps 0;

Matches 341; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VGRGDAILITCAVHPFLIDTKLLAIFGPLMLQAGITKVPYFVRAGLIRACMLV 60

Db 57 VGRGDAILITCAVHPFLIDTKLLAIFGPLMLQAGITKVPYFVRAGLIRACMLV 116

Qy 61 RKAAGHYYQMAFMKLAALTGTYYVDHUTPLQDWAHAGLIRDLAVALAVENIFSMEVKIT 120

Db 117 RKAAGHYYQMAFMKLAALTGTYYVDHUTPLQDWAHAGLIRDLAVALAVENIFSMEVKIT 176

Qy 121 WGADTAACGDIISGLPVSPSARRGREILGADNFFGQRLLAPITASQQTRGILGCIT 180

Db 177 WGADTAACGDIISGLPVSPSARRGREILGADNFFGQRLLAPITASQQTRGILGCIT 236

Qy 181 SLTGRDKNOVQEGETQVSTATOSFLATCNGVCMVTFEGAGSKTLAGPKPITQMYTNVD 240

Db 237 SLTGRDKNOVQEGETQVSTATOSFLATCNGVCMVTFEGAGSKTLAGPKPITQMYTNVD 296

Qy 241 QDLVGWQAPPGARSMTPTCGSSDLVYRHDVIPVRRRGDRGSLSPPRPSYLGKSS 300

Db 297 QDLVGWQAPPGARSMTPTCGSSDLVYRHDVIPVRRRGDRGSLSPPRPSYLGKSS 356

Qy 301 GGPILCPSPGHAVGIFRAAVCTRGYAKAVIPIVPMETM 341

Db 357 GGPILCPSPGHAVGIFRAAVCTRGYAKAVDPIPVESETM 397

RESULT 6  
 AAR82694 ID AAR82694 standard; protein; 3010 AA.

AC AAR82694;  
 XX DT 14-OCT-1996 (revised)  
 XX DT 14-NOV-1996 (first entry)

DE Partial HCV non-structural polyprotein.  
 KW proteinase; hepatitis C virus; screening; inhibitor; proteolytic;  
 KW identification; cleavage.  
 XX OS Hepatitis C virus.

XX FH Key Location/Qualifiers  
 FT Protein 898..1233 /note= "partial proteinase; see AAR82692"  
 FT Protein 922..1907 /note= "partial proteinase; see AAR82693"

XX PN JP07184648-A.  
 XX DD 25-JUL-1995.

XX PP 05-FEB-1993; 93JP-00018854.

XX PR 07-FEB-1992; 92JP-00022657.

PR 18-SEP-1992; 92JP-00249240.

PR 04-DEC-1992; 92JP-00325303.

(KAEN/) KARENNO K.  
 PA (SUMO ) SUMITOMO METAL IND LTD.

PA (SOYA-) SOYAKU GIJUTSU KENKYUSHO KK.

XX WPI; 1995-287962/38.

DR N-PSDB; AAT03960.

PT An HCV proteinase active substance - which has activity as an anti-HCV agent and can be used to screen for proteinase inhibitors.  
 PR XX Disclosure, Page 39-48; 52pp; Japanese.

PS The present sequence is a partial Hepatitis C Virus (HCV) polyprotein from the non-structural region. Partial proteinase sequences (AAR82692-93) are contained within this sequence. The proteinases can be used as anti-HCV agents. They can also be used to screen cpds. for their ability to inhibit their proteolytic activity. In this way proteinase inhibitors can be identified. (Updated on 16-OCT-2003 to standardise OS field)

SQ Sequence 3010 AA;

Query Match 96.2%; Score 1710; DB 2; Length 3010;

Best Local Similarity 94.7%; Pred. No. 8.3e-156; Mismatches 7; Indels 0; Gaps 0;

Matches 323; Conservative 11; Mismatches 7; Indels 0; Gaps 0;

Qy 1 VGRGDAILITCAVHPFLIDTKLLAIFGPLMLQAGITKVPYFVRAGLIRACMLV 60

Db 866 VGRGDAILITCAVHPFLIDTKLLAIFGPLMLQAGITKVPYFVRAGLIRACMLV 925

Qy 61 RKAAGHYYQMAFMKLAALTGTYYVDHUTPLQDWAHAGLIRDLAVALAVENIFSMEVKIT 120

Db 926 RKAAGHYYQMAFMKLAALTGTYYVDHUTPLQDWAHAGLIRDLAVALAVENIFSMEVKIT 985

Qy 121 WGADTAACGDIISGLPVSPSARRGREILGADNFFGQRLLAPITASQQTRGILGCIT 180

Db 986 WGADTAACGDIISGLPVSPSARRGREILGADNFFGQRLLAPITASQQTRGILGCIT 1045

Qy 181 SLTGRDKNOVQEGETQVSTATOSFLATCNGVCMVTFEGAGSKTLAGPKPITQMYTNVD 240

Db 1046 SLTGRDKNOVQEGETQVSTATOSFLATCNGVCMVTFEGAGSKTLAGPKPITQMYTNVD 1105

Qy 241 QDLVGWQAPPGARSMTPTCGSSDLVYRHDVIPVRRRGSLSPPRPSYLGKSS 300

Db 1106 QDLVGWQAPPGARSMTPTCGSSDLVYRHDVIPVRRRGSLSPPRPSYLGKSS 1165

Qy 301 GGPILCPSPGHAVGIFRAAVCTRGYAKAVDPIPVESETM 341

Db 1166 GGPILCPSPGHAVGIFRAAVCTRGYAKAVDPIPVESETM 397

RESULT 7  
 AAR68622 ID AAR68622 standard; protein; 3010 AA.

AC AAR68622;  
 XX DT 15-OCT-2003 (revised)  
 XX DT 16-OCT-1995 (first entry)

DE HCV protein cleavable with new serine proteinase.

XX OS proteinase; serine; cleavage; hepatitis C virus; HCV.

XX OS Hepatitis C virus; Virus.

XX FH Key Location/Qualifiers  
 FT Cleavage-site 2419..2420  
 FT /note= "Serine protease cleavage site"

XX PN JP06315377-A.

XX DD 15-NOV-1994.

XX PR 06-MAY-1993; 93JP-00105666.

XX PR 06-MAY-1993; 93JP-00105666.

XX PA (KAEN/) KARENNO K.  
 PA (SUMO ) SUMITOMO METAL IND LTD.

PA (SOYA-) SOYAKU GIJUTSU KENKYUSHO KK.

|   |  |  |  |  |
|---|--|--|--|--|
| XX  | DR   | FT   | Region   | 325..327   |
| DR  | N-PSDB;  | FT   | /label= N-linked glycosylation site  | 327..327   |
| XX  | AQ80498.   | FT   | 417..419   | 417..419   |
| PT  | New HCV-originated proteinase active substance - used for site-specific cleavage by an intermolecular reaction and the purification thereof.                     | FT   | 423..425   | 423..425   |
| XX  | Disclosure; Page 10-19; 23pp; Japanese.  | FT   | 430..432   | 430..432   |
| PS  | Sequence 3010 AA;  | FT   | 448..450   | 448..450   |
| XX  | Query Match 95.9%; Score 1705; DB 2; Length 3010; Best local Similarity 94.4%; Pred. No. 2.5e-155; Matches 322; Conservative 11; Mismatches 8; Indels 0; Gaps 0; | FT   | /label= N-linked glycosylation site  | 532..534   |
| QY  | 1 VRGGDAAILLTCAVHPPELFIDIKLLAIFGPMVLQAGITKPYFVRAGLIRACMV   | FT   | 532..534   | 532..534   |
| Db  | 866 VRGGDAAILLTCAVHPPELFIDIKLLAIFGPMVLQAGITRVPYFVRAGLIRACMV  | FT   | /label= N-linked glycosylation site  | 556..558   |
| QY  | 61 RKAAGHYYVQAFMKLAALTGTYYDHLPQLQDWAHAGRLDAAVEPVFSNEVKIT   | FT   | 556..558   | 556..558   |
| Db  | 926 RKVAGHYYVQAFMKLAALTGTYYDHLPQLQDWAHAGRLDAAVEPVFSNEVKIT  | FT   | /label= N-linked glycosylation site  | 623..625   |
| QY  | 121 WGADTAAGDPISGLPVSARRGRREITLGPGDADNFEGQGRLLAPITAYSOOTRGIGCIT  | FT   | 623..625   | 623..625   |
| Db  | 986 WGADTAAGDPISGLPVSARRGRREITLGPGDADNFEGQGRLLAPITAYSOOTRGIGCIT  | FT   | /label= N-linked glycosylation site  | 645..647   |
| QY  | 181 SITGRDKNQEGEQVNSTATQSLATCNGVCTVPHQGSKTLAGKGPTQMYNTD  | FT   | 645..647   | 645..647   |
| Db  | 1046 SITGRDKNQEGEVQVLSTHATQSLATCNGVCTVPHQGSKTLAGKGPTQMYNTD   | FT   | /label= N-linked glycosylation site  | 1213..1215   |
| QY  | 241 QDLVHQAPGARSMPCTCGSSDILVWHAUDIPVRRGDRGSLSPRPSYIKQSS  | FT   | 1213..1215   | 1213..1215   |
| Db  | 1106 QDLVHQAPGARSMPCTCGSSDILVWHAUDIPVRRGDRGSLSPRPSYIKQSS   | FT   | /label= N-linked glycosylation site  | 1255..1257   |
| QY  | 301 GGPLICPSGHAVGTFRAAVTRGVAKAVDFFPVVESMETMR   | FT   | 1255..1257   | 1255..1257   |
| Db  | 1166 GGPLICPSGHAVGTFRAAVTRGVAKAVDFFPVVESMETMR  | FT   | /label= N-linked glycosylation site  | 2340..2342   |
| RESULT 8  | PR 18-SEP-1992; 92JP-00249241.   | PD   | 22-NOV-1994.   | 22-NOV-1994.   |
| ID AAR68864   | XX   | PR   | 18-SEP-1992; 92JP-00249241.  | 18-SEP-1992;   |
| XX  | XX   | PA   | (SOYA-) SOYAKU GIUTSU KENKYUSHO KK.  | (SOYA-) SOYAKU GIUTSU KENKYUSHO KK.  |
| AC 06-DBC-1995  | XX   | DR   | WPI; 1995-040301/06.   | WPI; 1995-040301/06.   |
| DT (first entry)  | XX   | DR   | N-PSDB; AQ81559.   | N-PSDB; AQ81559.   |
| XX  | XX   | FT   | of hepatitis C virus helicase gene in baculovirus - useful for large scale prodn. of RNA helicase.   | of hepatitis C virus helicase gene in baculovirus - useful for large scale prodn. of RNA helicase.   |
| OS Hepatitis C virus.   | XX   | PS   | Claim 1; Fig 1-4; 9pp; Japanese.   | Claim 1; Fig 1-4; 9pp; Japanese.   |
| DE Hepatitis C virus RNA helicase.  | XX   | CC   | AQ81559 encodes Aar68864 hepatitis C virus (HCV) RNA helicase. The DNA was used in the construction of an expression vector, which was used to transform a baculovirus host. The transformed baculovirus could then be used for the recombinant prodn. of HCV RNA helicase | AQ81559 encodes Aar68864 hepatitis C virus (HCV) RNA helicase. The DNA was used in the construction of an expression vector, which was used to transform a baculovirus host. The transformed baculovirus could then be used for the recombinant prodn. of HCV RNA helicase |
| XX  | XX   | FT   | Sequence 3010 AA;  | Sequence 3010 AA;  |
| KW Hepatitis C virus; HCV; non-A non-B; helicase gene; RNA helicase; baculovirus; recombinant production. | XX   | Query Match 95.9%; Score 1704; DB 2; Length 3010; Best Local Similarity 94.4%; Pred. No. 3.2e-155; Matches 322; Conservative 11; Mismatches 8; Indels 0; Gaps 0; | Query Match 95.9%; Score 1704; DB 2; Length 3010; Best Local Similarity 94.4%; Pred. No. 3.2e-155; Matches 322; Conservative 11; Mismatches 8; Indels 0; Gaps 0;   |  |
| OS Hepatitis C virus.   | XX   | QY 1 VRGGDAAILLTCAVHPPELFIDIKLLAIFGPMVLQAGITKPYFVRAGLIRACMV  | QY 1 VRGGDAAILLTCAVHPPELFIDIKLLAIFGPMVLQAGITKPYFVRAGLIRACMV  |  |
| XX  | FT   | Db 866 VRGGDAAILLTCAVHPPELFIDIKLLAIFGPMVLQAGITRVPYFVRAGLIRACMV   | Db 866 VRGGDAAILLTCAVHPPELFIDIKLLAIFGPMVLQAGITRVPYFVRAGLIRACMV   |  |
| Key Location/Qualifiers   | XX   | QY 61 RKAAGHYYVQAFMKLAALTGTYYDHLPQLQDWAHAGRLDAAVEPVFSNEVKIT  | QY 61 RKAAGHYYVQAFMKLAALTGTYYDHLPQLQDWAHAGRLDAAVEPVFSNEVKIT  |  |
| FT Region   | 209..211   | Db 926 RKVAGHYYVQAFMKLAALTGTYYDHLPQLQDWAHAGRLDAAVEPVFSNEVKIT   | Db 926 RKVAGHYYVQAFMKLAALTGTYYDHLPQLQDWAHAGRLDAAVEPVFSNEVKIT   |  |
| FT Region   | 234..236   | QY 121 WGADTAAGDPISGLPVSARRGRREITLGPGDADNFEGQGRLLAPITAYSOOTRGIGCIT   | QY 121 WGADTAAGDPISGLPVSARRGRREITLGPGDADNFEGQGRLLAPITAYSOOTRGIGCIT   |  |
| FT Region   | 250..252   | Db 986 WGADTAAGDPISGLPVSARRGRREITLGPGDADNFEGQGRLLAPITAYSOOTRGIGCIT   | Db 986 WGADTAAGDPISGLPVSARRGRREITLGPGDADNFEGQGRLLAPITAYSOOTRGIGCIT   |  |
| FT Region   | 305..307   | QY 986 WGADTAAGDPISGLPVSARRGRREITLGPGDADNFEGQGRLLAPITAYSOOTRGIGCIT   | QY 986 WGADTAAGDPISGLPVSARRGRREITLGPGDADNFEGQGRLLAPITAYSOOTRGIGCIT   |  |
| FT Region   | 305..307   | Db 986 WGADTAAGDPISGLPVSARRGRREITLGPGDADNFEGQGRLLAPITAYSOOTRGIGCIT   | Db 986 WGADTAAGDPISGLPVSARRGRREITLGPGDADNFEGQGRLLAPITAYSOOTRGIGCIT   |  |
| FT Region   | 305..307   | QY 986 WGADTAAGDPISGLPVSARRGRREITLGPGDADNFEGQGRLLAPITAYSOOTRGIGCIT   | QY 986 WGADTAAGDPISGLPVSARRGRREITLGPGDADNFEGQGRLLAPITAYSOOTRGIGCIT   |  |

CC the claims of the invention  
 XX Sequence 2201 AA;  
 SQ

Query Match 95 %; Score 1722; DB 5; Length 2201;  
 Best Local Similarity 94.1%; Pred. No. 3.2e-155; Mismatches 9; Indels 0; Gaps 0;

Matches 321; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

Db 1106 QDLVGVWAPPGRASMTPTCGSSDLYLVRHADVIPVRRGSRGSLLSPRVSYIKGS 300  
 Qy 301 GGPLCPSGHAVGIFRAAVCTRGVAKAVDIPVEMETMR 341  
 Db 1166 GGPLCPSGHVGIFRAAVCTRGVAKAVDIPVEMETMR 1206

RESULT 9  
 ABQ30601  
 ID ABQ30601 standard; protein; 2201 AA.  
 XX ABQ30601;  
 XX 21-OCT-2002 (first entry)  
 DT Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B mutant #10.  
 XX Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;  
 KW cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; mutein.  
 XX CS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 882  
 FT /label= Arg, Lys  
 FT  
 FT Misc-difference 2183  
 FT /note= "Wild type Met substituted by Thr"  
 XX WO200252015-A2.  
 XX  
 PD 04-JUL-2002.  
 XX  
 PP 20-DEC-2001; 2001WO-CA001843.  
 XX PR 22-DEC-2000; 2000US-0257857P.  
 XX PA (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  
 XX PT Kukolj, G., Pause, A.;  
 PT XX  
 DR WPI; 2002-575382/61.  
 XX  
 PT New self-replicating RNA molecules from Hepatitis C virus (HCV), which  
 PT possess enhanced transduction or replication efficiency, useful for  
 PT evaluating potential inhibitors of HCV replication.  
 XX  
 PS Claim 3; Page: 140PP; English.

XX The invention describes a self-replicating hepatitis C virus (HCV) polyribonucleotide molecule comprising a 5'-non translated region (NTR), where quanine at position 1 is substituted for adenine, a HCV polyprotein region coding for a HCV polyprotein, and a 3'-NTR region. The self-replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating potential inhibitors of HCV replication. The HCV RNA molecule is also useful for efficiently establishing cell culture replication. The self-replicating polyribonucleotide molecule contains a 5'-NTR, where G at position 1 is substituted for A, and therefore provides an alternative to existing systems comprising a self-replicating HCV RNA molecule that, in conjunction with mutations in the HCV non structural region, such as the G(2042)C/R mutations, transduces and/or replicates with greater efficiency. This amino acid sequence represents a mutant of the hepatitis C virus replicon APGXL2 and contains the wild protease NS2/3 protease complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note: this sequence does not appear in the specification but has been created from the wild type sequence shown in ABG30580 using information given in

CC  
 XX  
 SQ

Query Match 95 %; Score 1722; DB 5; Length 2201;  
 Best Local Similarity 94.1%; Pred. No. 3.2e-155; Mismatches 9; Indels 0; Gaps 0;

Matches 321; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

Db 57 VRGGRDAILITCAHPELIFPTINKILLATLGPMVNLQAGITKPYFVRQGLIRACMV 60  
 Qy 61 RKAAGCHYVQNAFMKLAALGTGYVDHPLQDWAHAGRDIAVAPVIFSDMEVKIT 120  
 Db 117 RKVAGCHYVQNAFMKLAALGTGYVDHPLQDWAHAGRDIAVAPVIFSDMEVKIT 175

Db 121 WGADTAACGDTISGPGVSAARGRGEGTLLGRADNFSGQWRLLAPLTAYSOPTRGILGICIT 180  
 Db 177 WGADTAACGDTISGPGVSAARGRGEGTLLGRADNFSGQWRLLAPLTAYSOPTRGILGICIT 236

Qy 181 SLTGRDKNQVEGVQVYSTATOSFLATCNGVCMVYHAGSKTLAGPKPITOMTN 240  
 Db 237 SLTGRDRNQVEGVQVYSTATOSFLATCNGVCMVYHAGSKTLAGPKPITOMTN 296

Qy 241 QDLVGVWAPPGRASMTPTCGSSDLYLVRHADVIPVRRGSRGSLLSPRVSYIKGS 300  
 Db 297 QDLVGVWAPPGRASMTPTCGSSDLYLVRHADVIPVRRGSRGSLLSPRVSYIKGS 356

Qy 301 GGPLCPSGHAVGIFRAAVCTRGVAKAVDIPVEMETMR 341  
 Db 357 GGPLCPSGHAVGIFRAAVCTRGVAKAVDIPVEMETMR 397

RESULT 10  
 ABG3091  
 ID ABG3091 standard; protein; 2201 AA.  
 XX ABG3091;  
 XX 21-OCT-2002 (first entry)  
 DT Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B mutant #3.  
 XX Self replicating; hepatitis C virus; HCV; HCV replication inhibitor;  
 KW cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; mutein.  
 XX CS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 751  
 FT /note= "Wild type Ser substituted by Gly"  
 FT Misc-difference 882  
 FT /label= Arg, Lys  
 XX WO200252015-A2.  
 XX  
 PD 04-JUL-2002.  
 XX PR 20-DEC-2001; 2001WO-CA001843.  
 XX PR 22-DEC-2000; 2000US-0257857P.  
 XX PA (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  
 XX PT Kukolj, G., Pause, A;  
 XX DR WPI; 2002-575382/61.  
 XX  
 PT New self-replicating RNA molecules from Hepatitis C virus (HCV), which  
 PT possess enhanced transduction or replication efficiency, useful for  
 PT evaluating potential inhibitors of HCV replication.



|                       |   |                             |    |   |
|-----------------------|---|-----------------------------|----|---|
| ID                    | ABG30581  | standard; protein; 2201 AA. | Db | 297 QDLVGSQAPPGARSITPCTCGSSDLYLVTRHADVLPVRGRGDSRGSLISPRPVSYLKGS 356 |
| XX                    |   |                             | Qy | 301 GGPLCPSGHAVGIRAAVCTRGVAKAVDIPVSMETMR 341                        |
| AC                    |   |                             | ID | ABG30593  |
| XX                    | 21-OCT-2002   | (first entry)               | XX | 357 GGPLCPSGHAVGIFRAAVCTRGVAKAVDIPVSMETMR 397                       |
| DT                    |   |                             | Ds |   |
| XX                    | Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B #1.  |                             |    |   |
| DB                    |   |                             |    |   |
| XX                    | Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;  |                             |    |   |
| KW                    | cell culture replication; NS2/3; NS3/4; NS3; NS5B.  |                             |    |   |
| OS                    |   |                             |    |   |
| XX                    | Hepatitis C virus.  |                             |    |   |
| PN                    | WO200252015-A2.   |                             |    |   |
| XX                    | 04-JUL-2002.  |                             |    |   |
| PD                    |   |                             |    |   |
| XX                    | 20-DEC-2001; 2001WO-CA001843.   |                             |    |   |
| PF                    |   |                             |    |   |
| XX                    | 22-DEC-2000; 2000US-0257857P.   |                             |    |   |
| PR                    |   |                             |    |   |
| XX                    | (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  |                             |    |   |
| PA                    |   |                             |    |   |
| XX                    | Kukolj, G., Pause A;  |                             |    |   |
| PI                    |   |                             |    |   |
| XX                    | DR  |                             |    |   |
| DR                    | N-PSDB; ABK88573.   |                             |    |   |
| XX                    | WPI; 2002-575382/61.  |                             |    |   |
| PT                    |   |                             |    |   |
| PT                    | New self-replicating RNA molecules from Hepatitis C virus (HCV), which  |                             |    |   |
| PT                    | possess enhanced transduction or replication efficiency, useful for evaluating potential inhibitors of HCV replication.   |                             |    |   |
| XX                    |   |                             |    |   |
| PS                    | Disclosure; Page 49-58; 140pp; English.   |                             |    |   |
| XX                    |   |                             |    |   |
| CC                    | The invention describes a self-replicating hepatitis C virus (HCV) polynucleotide molecule comprising a 5'-non translated region (NTR), where guanine at position 1 is substituted for adenine, a HCV polyprotein region coding for a HCV polyprotein; and a 3'-NTR region. The self-replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating potential inhibitors of HCV replication. The HCV RNA molecule is also useful for efficiently establishing cell culture replication. The molecule is also replicating polynucleotide molecule contains a 5'-NTR, where G at position 1 is substituted for A, and therefore provides an alternative to existing systems comprising a self-replicating HCV RNA molecule that, in conjunction with mutations in the HCV non-structural region, such as the G(2042)C/R mutations, transduces and/or replicates with greater efficiency. This amino acid sequence represents a mutant of the hepatitis C virus replicon APGK12 and contains the viral protease NS2/3, protease CC NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. |                             |    |   |
| CC                    | replicon APGK12 and contains the viral protease NS2/3, protease complex CC NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B.   |                             |    |   |
| XX                    |   |                             |    |   |
| SQ                    | Sequence 2201 AA;   |                             |    |   |
| Query Match           | 95.7%; Score 1702; DB 5; Length 2201;   |                             |    |   |
| Best Local Similarity | 94.1%; Pred. No. 3.2e-155;  |                             |    |   |
| Matches               | 321; Conservative 11; Mismatches 9; Indels 0; Gaps 0;   |                             |    |   |
| Qy                    |   |                             |    |   |
| 1                     | VRGGRDAATLTLTCVPHPELIFPDITKLIAIFGPIMVLRGKIKYPPYFVRAGSIIRACMLV 60  |                             |    |   |
| Db                    | 57 VRRGGDAVILTCIHPPELIFPDITKLIAIFGPIMVLRGKIKYPPYFVRAGSIIRACMLV 116  |                             |    |   |
| Qy                    | 61 RKAAGHYWQAFMKIAALTGTYYDHTPLQWAHAGRDLLAWAVRVIIFPSMEVKIT 120   |                             |    |   |
| Db                    | 117 RVKAGHYWQAMKIAALTGTYYDHTPLQWAHAGRDLLAWAVRVIIFPSMEVKIT 176   |                             |    |   |
| Qy                    | 121 WGATAAACDDIISGLPVSARRGRILGILGPMNFGEGQRILLAPITAQSQTGIGSIT 180  |                             |    |   |
| Db                    | 177 WGADTAACGDIILGLPVSARRGRILGILGPMNFGEGQRILLAPITAQSQTGIGSIT 236  |                             |    |   |
| Qy                    | 181 SLMTRDKNQYEVGEYQVVSATQSTLACTCYNGVNTVEFAGSKTLASIGKGPFTQMYIVD 240   |                             |    |   |
| Db                    | 237 SLTGRDRNQYEVGEYQVVSATQSTLACTCYNGVNTVEFAGSKTLASIGKGPFTQMYIVD 296   |                             |    |   |
| Qy                    | 241 QDLVGSQAPPGARSITPCTCGSSDLYLVTRHADVLPVRGRGDSRGSLISPRPVSYLKGS 300   |                             |    |   |
| Query Match           | 95.7%; Score 1702; DB 5; Length 2201;   |                             |    |   |
| Best Local Similarity | 94.1%; Pred. No. 3.2e-155;  |                             |    |   |

Matches 321; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

QY 1 VRGGDAILITCAVHPPELFIDIKLIAIFGDLMLQAGITKPYFRAQGLIRACLV 60  
Db 57 VRGGDAILITCAIHPPELIFTIKLIALGPMVLQAGITKPYFRAQGLIRACLV 116

QY 61 RKAAGHHYVQMAFMKLAULTGTYVDHPLQDWAHAGRLDAVAVPVISFMEVKIT 120  
Db 117 RKVAGHHYVQMAFMKLAULTGTYVDHPLQDWAHAGRLDAVAVPVISFMEVKIT 176

QY 121 WGADTAACGDTISGLPVSAARGRETLIGPDAFNFGQGRLLAPITAYSQTRGLGCT 180  
Db 177 WGADTAACGDTISGLPVSAARGRETLIGPDAFNFGQGRLLAPITAYSQTRGLGCT 236

QY 181 SLTGRDNQVEGEVQVSTATOSFLATCNGVCWTFVFGASKTLAGPKGPIYOMYIVND 240  
Db 237 SLTGRDNQVEGEVQVSTATOSFLATCNGVCWTFVFGASKTLAGPKGPIYOMYIVND 296

QY 241 QDLVGWQAPPGARSMTPTCGSSDLYLVRHADWIPVRRGDSRSLSPPRVSYLGSS 300  
Db 297 QDLVGWQAPPGARSLTPTCGSSDLYLVRHADWIPVRRGDSRSLSPPRVSYLGSS 356

QY 301 GGPLLCPSGHAVGFRAVCTRGVAKAVDIPFVPSMETM 341  
Db 357 GGPLLCPSGHAVGFRAVCTRGVAKAVDIPFVPSMETM 397

RESULT 14

ID ABG30582  
ID ABG30582 standard; protein: 2201 AA.

AC ABG30582;  
XX DT 21-OCT-2002 (first entry)

XX DE Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B #2.

XX KW Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor; cell culture replication; NS2/3; NS3/4; NS3; NS5B; mutant; mutain.

XX OS Hepatitis C virus.

XX FH Location/Qualifiers

PT Misc-difference 882  
PT /note= "Wild type Lys substituted by Lys or Arg"  
PT /note= "Wild type Gly substituted by Cys"  
XX PN WO200252015-A2.

PD 04-JUL-2002.

XX PF 20-DEC-2001; 2001WO-CA001843.

XX PR 22-DEC-2000; 2000US-0257857P.

XX PA (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.

XX PI Kukolj G, Pause A,  
XX DR WPI: 2002-573382/61..  
XX DR N-ISDB; ABK85574.

XX PT New self-replicating RNA molecules from Hepatitis C virus (HCV), which possess enhanced transduction or replication efficiency, useful for evaluating potential inhibitors of HCV replication.

XX PS Disclosure; Page 59-69; 140pp; English.

CC CC The invention describes a self-replicating hepatitis C virus (HCV) polynucleotide molecule comprising a 5'-non translated region (NTR), where guanine at position 1 is substituted for adenine, a HCV polyprotein region coding for a HCV polyprotein; and a 3'-NTR region. The self-

CC CC replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating potential inhibitors of HCV replication. The HCV RNA molecule is also useful for efficiently establishing cell culture replication. The self-replicating polynucleotide molecule contains a 5'-NTR, where G at position 1 is substituted for A, and therefore provides an alternative to existing systems comprising a self-replicating HCV RNA molecule that, in conjunction with mutations in the HCV non-structural region, such as the G(2042)C/R mutations, transduces and/or replicates with greater efficiency. This amino acid sequence is encoded by the hepatitis C virus replicon APGK12 and contains the viral protease NS2/3, protease complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NS5B. Note: this sequence has been created from replicon APGK12 shown in ABG30581

XX SQ Sequence 2201 AA;

Query Match 95.7%; Score 1702; DB 5; Length 2201;  
Best Local Similarity 94.1%; Pred. No. 3.2e-155; Indels 0; Gaps 0;  
Matches 321; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

QY 1 VRGGDAILITCAVHPPELFIDIKLIAIFGDLMLQAGITKPYFRAQGLIRACLV 60  
Db 57 VRGGDAILITCAIHPPELIFTIKLIALGPMVLQAGITKPYFRAQGLIRACLV 116

QY 61 RKAAGHHYVQMAFMKLAULTGTYVDHPLQDWAHAGRLDAVAVPVISFMEVKIT 120  
Db 117 RKVAGHHYVQMAFMKLAULTGTYVDHPLQDWAHAGRLDAVAVPVISFMEVKIT 176

QY 121 WGADTAACGDTISGLPVSAARGRETLIGPDAFNFGQGRLLAPITAYSQTRGLGCT 180  
Db 177 WGADTAACGDTISGLPVSAARGRETLIGPDAFNFGQGRLLAPITAYSQTRGLGCT 236

QY 181 SLTGRDNQVEGEVQVSTATOSFLATCNGVCWTFVFGASKTLAGPKGPIYOMYIVND 240  
Db 237 SLTGRDNQVEGEVQVSTATOSFLATCNGVCWTFVFGASKTLAGPKGPIYOMYIVND 296

QY 241 QDLVGWQAPPGARSMTPTCGSSDLYLVRHADWIPVRRGDSRSLSPPRVSYLGSS 300  
Db 297 QDLVGWQAPPGARSLTPTCGSSDLYLVRHADWIPVRRGDSRSLSPPRVSYLGSS 356

QY 301 GGPLLCPSGHAVGFRAVCTRGVAKAVDIPFVPSMETM 341  
Db 357 GGPLLCPSGHAVGFRAVCTRGVAKAVDIPFVPSMETM 397

RESULT 15

ID ABG30580  
ID ABG30580 standard; protein: 2201 AA.

AC ABG30580;  
XX DT 21-OCT-2002 (first entry)

XX DE Hepatitis C virus NS2/3, NS3/4, NS3 and NS5B #9.

XX KW Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor; cell culture replication; NS2/3; NS3/4; NS3; NS5B.

XX OS Hepatitis C virus.

XX FH Location/Qualifiers

PT Misc-difference 882  
PT /note= "Encoded by ARG"  
XX PN WO200252015-A2.  
XX PD 04-JUL-2002.  
XX PF 20-DEC-2001; 2001WO-CA001843.  
XX PR 22-DEC-2000; 2000US-0257857P.  
XX PA (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.

PI Kukolj G, Pause A;  
 XX  
 DR WPI; 2002-575382/61.  
 XX

PT New self-replicating RNA molecules from Hepatitis C virus (HCV), which possess enhanced transduction or replication efficiency, useful for evaluating potential inhibitors of HCV replication.  
 XX

PS Disclosure; Page 69-74; 140pp; English.

XX  
 CC The invention describes a self-replicating hepatitis C virus (HCV) poly nucleotide molecule comprising a 5'-non translated region (NTR), where guanine at position 1 is substituted for adenine, a HCV polyprotein region coding for Hepatitis C Virus (HCV) RNA molecule is useful for evaluating potential inhibitors of HCV replication. The HCV RNA molecule is also replicating poly nucleotide molecule contains a 5'-NTR, where G at position 1 is substituted for A, and therefore provides an alternative to existing systems comprising a self-replicating HCV RNA molecule that, in conjunction with mutations in the HCV non-structural region, such as the G(2042)C/R mutations, transduces and/or replicates with greater efficiency. This amino acid sequence is encoded by the hepatitis C virus replicon APGK12 and contains the viral protease NS2/3, protease complex CC NS3/4, helicase NS3 and RNA-dependent RNA polymerase NSB  
 XX

Sequence 2201 AA;

SQ Query Match 95.7%; Score 1702; DB 5; Length 2201;  
 Best Local Similarity 94.1%; Pred. No. 3.2e-155;  
 Matches 321; Conservative 11; Mismatches 9; Indels 0; Gaps 0;  
 QY 1 VRGERDAITLTCVHPELIFPDTRKLLAIFGPMVHQAGIRKTPYFRAQQSRIRALV 60  
 Db 57 VRRGRDAVILTCVHPELIFPDTRKLLAIFGPMVHQAGIRKTPYFRAHQGLIRACMLV 116  
 QY 61 RKAAGGHHYQMAFMVKLAALTGTGVYDHJPLQWHAQSLRDAVAVPVIDMEVKIT 120  
 Db 117 RKVAGGHYQMAFMVKLAALTGTGVYDHJPLQWHAQSLRDAVAVPVIDMEVKIT 176  
 QY 121 WGDTAAAGDDIISGLPVSVSARRGRERLIGPADNGEQQRLLAPITAYSQQTGILGIT 180  
 Db 177 WGDTAAAGDDIISGLPVSVSARRGRERLIGPADNGEQQRLLAPITAYSQQTGILGIT 236  
 QY 181 SLDGRDRNQVEGEYQVVSATQSPLATCNGVWTQVHGAGSKTLAEGKPTQMVND 240  
 Db 237 SLTGRDRNQVEGEYQVVSATQSPLATCNGVWTQVHGAGSKTLAEGKPTQMVND 296  
 QY 241 QDLYGWQAPPGARSNTPTCGGSSDILYVTRHADYIPVRGRGSRSLSSPRPSYLGSS 300  
 Db 297 QDLYGWQAPPGARSNTPTCGGSSDILYVTRHADYIPVRGRGSRSLSSPRPSYLGSS 356  
 QY 301 GGRPLCPSPHAGVGFRAAVCTRGVAKAVDFPIVYESMETMR 341  
 Db 357 GPLLCPSGHAGVGFRAAVCTRGVAKAVDFPVYESMETMR 397



OM protein - protein search, using sw model

Run on: May 6, 2004, 09:22:36 ; Search time 11.1056 Seconds  
 (without alignments)  
 2913.573 Miligion cell updates/sec

Title: US-10-650-585-14  
 Perfect score: 1778  
 Sequence: VRGGRDAAILLTCAVHPMLI.....RGVAKAVDPIPVSMEETMR 341  
 Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext: 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
 Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries

Database : PIR 78:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

**RESULT 1**

A45573

genome polyprotein - hepatitis C virus (strain JT)

Nr: contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructural protein NS4a); nonstructural protein NS4b; nonstructural protein NS5

C: Species: hepatitis C virus

C: Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 19-Jan-2001

C: Accession: A45573

R:Tanaka, T.; Kato, N.; Nakagawa, M.; Ootsuyama, Y.; Cho, M.J.; Nakazawa, T.; Hijikata, A.; Virus Res., 23, 35-53, 1992

A: Title: Molecular cloning of hepatitis C virus genome from a single Japanese carrier: a molecular approach

A: Reference number: A45573; MUID:92295714; PMID:1318627

A: Accession: A45573

A: Status: preliminary

A: Molecule type: DNA

A: Residues: 1-3010 <PAX>

A: Cross-references: GB:D1168; GB:D01171; NID:9221612; PID:BA01943.1; PID:9221613

A: Experimental source: HCV-JT

A: Note: sequence extracted from NCBI backbone (NCBIN:106206, NCBIPI:106207)

C: Superfamily: hepatitis C virus genome polyprotein

C: Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serine/threonine kinase; structural protein

F: 2-15/Product: capsid protein C #status Predicted <CP>

F: 116-191/Product: envelope protein M #status Predicted <CP>

F: 192-389/Product: major envelope protein B #status Predicted <CP>

F: 390-729/Product: nonstructural protein NS1 #status Predicted <NS1>

F: 730-1006/Product: nonstructural protein NS2 #status Predicted <NS2>

F: 1007-1615/Product: hepacivirin #status Predicted <NS3>

F: 1231-1237/Region: nucleotide-binding motif A (P-loop)

F: 1310-1317/Region: nucleotide-binding motif B

F: 1318-1319/Region: DBXH motif

F: 1611-1862/Product: nonstructural protein NS4a #status Predicted <NS4>

F: 1863-2013/Product: nonstructural protein NS5b #status Predicted <NS5>

F: 2014-3010/Product: nonstructural protein NS5 #status Predicted <NS5>

Query Match 96.2%; Score 1711; DB 1; Length 3010;  
 Best Local Similarity 95.3%; Pred. No. 47e-136;  
 Matches 325; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

Qy VRGGRDAAILLTCAVHPMLI.....RGVAKAVDPIPVSMEETMR 60  
 Db VRGGRDAAILLTCAVHPMLI.....RGVAKAVDPIPVSMEETMR 925

Qy RRAAGGGYVQMFKAQALTTGTVYDHLPDWAAHLGRDLAVAVEPVIFDMEVKLT 120  
 Db RRAAGGGYVQMFKAQALTTGTVYDHLPDWAAHLGRDLAVAVEPVIFDMEVKLT 985

Qy WQADTAACGDIISGLPVASGRREILGPANFEGCQWRLAPIVATYSQQTRRGLGKLT 180  
 Db WQADTAACGDIISGLPVASGRREILGPANFEGCQWRLAPIVATYSQQTRRGLGKLT 1045

Qy SITGRDENQVGEVGQVSTAQFSFLATCNGCWTMAGGSKTLAGPKGKPTQMYWIND 240

|   |   |  |      |
|---|---|--|------|
| Db  | 1146  | SITGRDKNQVEGEVQVSTATOSFLATCNGVCHVFGAGSKTLAQPKGKPIOTMYTNVD  | 1105 |
| Qy  | 241   | QDLVGWQAPPGRASMTPTCGSSDLYLVTRHADVIPVRRGDSRGSSLSPRSVYLGSS   | 300  |
| Db  | 1106  | QDLVGWHPAPPGRASLTPCTCGSSDLYLVTRHADVIPVRRGDSRGSSLSPRSVYLGSS | 1165 |
| Qy  | 301   | GGLPLCPGSHAVGIFRAAVCTRGVAKAVDPFIPVESMETMR                  | 341  |
| Db  | 1166  | GGLPLCPGSHAVGIFRAAVCTRGVAKAVDPFIPVESMETMR                  | 1206 |
| RESULT 2  |   |  |      |
| GNWVJ   | genome polyprotein - hepatitis C virus (strain J)   |  |      |
| N;Contains:   | capsid protein C; envelope protein M; major envelope protein E; nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5     |  |      |
| C;Species:  | hepatitis C virus   |  |      |
| C;Accession:  | 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 19-Jan-2001   |  |      |
| A;Title:  | Molecular cloning of the human hepatitis C virus genome from Japanese patients  |  |      |
| A;Reference number:   | A39253; MUID:91088550; PMID:2175903   |  |      |
| A;Molecule type:  | genomic RNA   |  |      |
| A;Residues:   | 1-3010 <CAT>  |  |      |
| A;Cross-references:   | GB:D90208; NID:921610; PIDN:BAA14233.1; PID:9221611   |  |      |
| R;Kato, N.; Hijikata, M.; Ootsuyama, Y.; Nakagawa, M.; Ohkoshi, S.; Sugimura, T.; Shimotohno, K.                      |   |  |      |
| Proc. Jpn. Acad. 65B, 219-223, 1989   |   |  |      |
| A;Title:  | Japanese isolates of the non-A-, non-B hepatitis viral genome show sequence variation   |  |      |
| A;Reference number:   | FE0085  |  |      |
| A;Accession:  | PS0086  |  |      |
| A;Molecule type:  | genomic RNA   |  |      |
| A;Residues:   | 2650-2707 <KA2>   |  |      |
| A;Experimental source:  | Japanese isolate  |  |      |
| A;Comment:  | The cleavage sites of this polyprotein have not been determined.  |  |      |
| C;Superfamily:  | hepatitis C virus genome polyprotein  |  |      |
| C;Keywords:   | ATP; glycoprotein; hydrolyase; nucleotide binding; P-loop; polyprotein; serine  |  |      |
| F1-15/Product:  | capsid protein C #status predicted <CPC>  |  |      |
| F116-191/Product:   | envelope Protein M #status predicted <EPM>  |  |      |
| F192-389/Product:   | major envelope protein E #status predicted <ME>   |  |      |
| F390-729/Product:   | nonstructural protein NS1 #status predicted <NS1>   |  |      |
| F730-106/Product:   | nonstructural protein NS2 #status predicted <NS2>   |  |      |
| F1007-1615/Product:   | hepacivirin #status predicted <NS3>   |  |      |
| F1230-1237/Region:  | nucleotide-binding motif B (P-loop)   |  |      |
| F1312-1317/Region:  | nucleotide-binding motif A (P-loop)   |  |      |
| F1316-1319/Region:  | DEKH motif  |  |      |
| F1316-1319/Region:  | DEKH motif  |  |      |
| F1863-2013/Product:   | nonstructural protein NS4a #status predicted <NS4A>   |  |      |
| F1863-2013/Product:   | nonstructural protein NS4b #status predicted <NS4B>   |  |      |
| F196,209,233,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,2077,2240,2241,2014-3010/Product: | nonstructural protein NS5 #status predicted <NS5>   |  |      |
| Query Match   | 94.8%; Score 1685; DB 1; Length 3010; Best Local Similarity 92.7%; Pred. No. 5.7e-136; Matches 316; Conservative 14; Mismatches 11; Indels 0; Gaps 0; |  |      |
| Qy  | 1 VRGERDAITLTCAVHPENLFDITKLIAIFGPMLVQAGITKPYFVRAGLIRACMLV   | 60   |      |
| Db  | 866 VRGERDAITLTCAVHPENLFDITKLIAIFGPMLVQAGITKPYFVRAGLIRACMLV   | 925  |      |
| Qy  | 61 RKAAGHGYVOMAFNKLALTTGTYVDDHLPQLQDAHAGRLDAVAVEPVIFSDNEVKLT  | 120  |      |
| Db  | 926 RKAAGHGYVOMAFNKLALTTGTYVDDHLPQLQDAHAGRLDAVAVEPVIFSDNEVKLT   | 985  |      |
| Qy  | 121 WGATAAACDIISGPVSPARRGRELLIGADNFSGQWILLAPTAYSQOTRGIGCIT  | 180  |      |
| Db  | 986 WGATAAACDIISGPVSPARRGRELLIGADNFSGQWILLAPTAYSQOTRGIGCIT  | 1045   |      |
| Qy  | 1 VRGERDAITLTCAVHPENLFDITKLIAIFGPMLVQAGITKPYFVRAGLIRACMLV   | 60   |      |
| Db  | 866 VRGERDAITLTCAVHPENLFDITKLIAIFGPMLVQAGITKPYFVRAGLIRACMLV   | 925  |      |
| Qy  | 61 RKAAGHGYVOMAFNKLALTTGTYVDDHLPQLQDAHAGRLDAVAVEPVIFSDNEVKLT  | 120  |      |
| Db  | 926 RKAAGHGYVOMAFNKLALTTGTYVDDHLPQLQDAHAGRLDAVAVEPVIFSDNEVKLT   | 985  |      |
| Qy  | 181 SLTGRDKNQVEGEVQVSTATOSFLATCNGVCHVFGAGSKTLAQPKGKPIOTMYTNVD   | 240  |      |
| Db  | 1046 SLTGRDKNQVEGEVQVSTATOSFLATCNGVCHVFGAGSKTLAQPKGKPIOTMYTNVD  | 1105   |      |
| Qy  | 241 QDLVGWQAPPGRASMTPTCGSSDLYLVTRHADVIPVRGGDSRGSLLSPRSVYLGSS  | 300  |      |
| Db  | 1106 QDLVGWHPAPPGRASLTPCTCGSSDLYLVTRHADVIPVRGGDSRGSLLSPRSVYLGSS   | 1165   |      |
| Qy  | 301 GGLPLCPGSHAVGIFRAAVCTRGVAKAVDPFIPVESMETMR   | 341  |      |
| Db  | 1166 GGLPLCPGSHAVGIFRAAVCTRGVAKAVDPFIPVESMETMR  | 1206   |      |
| RESULT 3  |   |  |      |
| GNWVW   | genome polyprotein - hepatitis C virus (strain Taiwan)  |  |      |
| N;Contains:   | capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructural protein NS4a); nonstructural protein NS4b; nonstructural protein NS5  |  |      |
| C;Species:  | hepatitis C virus   |  |      |
| C;Note:   | host Homo sapiens (man)   |  |      |
| C;Date:   | 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 19-Jan-2001   |  |      |
| C;Accession:  | A40244  |  |      |
| R;Chen, P.J.; Lin, M.H.; Tsai, K.F.; Liu, P.C.; Lin, C.J.; Chen, D.S.   |   |  |      |
| Virology 189, 102-113, 1992   |   |  |      |
| A;Title:  | The Taiwanese hepatitis C virus genome: sequence determination and mapping to the genome of hepatitis C virus   |  |      |
| A;Reference number:   | A40244; MUID:92230206; PMID:1314449   |  |      |
| A;Accession:  | GB:MB4754   |  |      |
| A;Cross-references:   | hepatitis C virus genome polyprotein  |  |      |
| C;Superfamily:  | hepatitis C virus genome polyprotein  |  |      |
| C;Keywords:   | ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5     |  |      |
| F1-15/Product:  | capsid protein C #status predicted <CPC>  |  |      |
| F116-191/Product:   | envelope Protein M #status predicted <EPM>  |  |      |
| F192-389/Product:   | major envelope protein E #status predicted <ME>   |  |      |
| F390-729/Product:   | nonstructural protein NS1 #status predicted <NS1>   |  |      |
| F730-106/Product:   | nonstructural protein NS2 #status predicted <NS2>   |  |      |
| F1007-1615/Product:   | hepacivirin #status predicted <NS3>   |  |      |
| F1230-1237/Region:  | nucleotide-binding motif B (P-loop)   |  |      |
| F1312-1317/Region:  | nucleotide-binding motif A (P-loop)   |  |      |
| F1316-1319/Region:  | DEKH motif  |  |      |
| F1616-1862/Product:   | nonstructural protein NS4a #status predicted <NS4A>   |  |      |
| F1863-2013/Product:   | nonstructural protein NS5 #status predicted <NS5>   |  |      |
| F196,209,233,234,250,305,325,417,423,430,448,532,540,556,576,623,645,1213,1255,2041,2077,2240,2241,2014-3010/Product: | nonstructural protein NS5 #status predicted <NS5>   |  |      |
| Query Match   | 94.8%; Score 1685; DB 1; Length 3010; Best Local Similarity 92.7%; Pred. No. 7.5e-134; Matches 316; Conservative 14; Mismatches 11; Indels 0; Gaps 0; |  |      |
| Qy  | 1 VRGERDAITLTCAVHPENLFDITKLIAIFGPMLVQAGITKPYFVRAGLIRACMLV   | 60   |      |
| Db  | 866 VRGERDAITLTCAVHPENLFDITKLIAIFGPMLVQAGITKPYFVRAGLIRACMLV   | 925  |      |
| Qy  | 61 RKAAGHGYVOMAFNKLALTTGTYVDDHLPQLQDAHAGRLDAVAVEPVIFSDNEVKLT  | 120  |      |
| Db  | 926 RKAAGHGYVOMAFNKLALTTGTYVDDHLPQLQDAHAGRLDAVAVEPVIFSDNEVKLT   | 985  |      |
| Qy  | 121 WGATAAACDIISGPVSPARRGRELLIGADNFSGQWILLAPTAYSQOTRGIGCIT  | 180  |      |
| Db  | 986 WGATAAACDIISGPVSPARRGRELLIGADNFSGQWILLAPTAYSQOTRGIGCIT  | 1045   |      |
| RESULT 4  |   |  |      |
| S18030  | genome polyprotein - hepatitis C virus (isolate JK1)  |  |      |

N-Contains: capsid protein C; envelope protein M; hepatitis C virus protein NS4A; nonstructural protein NS4b; nonstructural protein NS5  
 C-Species: hepatitis C virus

A-Variety: isolate JK1  
 A-Accession: S18020  
 C-Accession: 19-May-2000 #sequence-revision 19-May-2000 #text\_change 23-Mar-2001  
 R-Honda, M.; Kaneko, S.; Masashi, U.; Kobayashi, K.; Murakami, S.  
 Submitted to the EMBL Data Library, September 1991  
 A-Description: A whole genome of hepatitis C virus cDNA was isolated from a single patient.  
 A-Reference number: A48332; MURD:93119270; PMID:8380322  
 A-Molecule type: genomic RNA  
 A-Residues: 1-3010 <HON>  
 A-Cross-references: EMBL:X61596; NID:959478; PIDN:CA43793.1; PID:959479  
 A-Experimental source: isolate JK1 from an individual  
 A-Arch. Viral. 128, 163-169, 1993  
 A-Title: Sequence analysis of putative structural regions of hepatitis C virus isolated  
 A-Accession: S33570  
 A-Cross-references: ENBIL:X61591  
 A-Note: this sequence is inconsistent with the nucleotide translation  
 A-Note: the authors translated the codon AGG for residue 43 as Pro, TCG for residue 320  
 A-Note: sequence extracted from NCBI backbone (NCBIN:121747, NCBI:121748)  
 C-Superfamily: hepatitis C virus genome polyprotein  
 C-Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serin  
 F-2-15/Product: capsid protein C #status predicted <CPC>  
 F-116-191/Product: envelope protein M #status predicted <EP>  
 F-120-389/Product: major envelope protein E #status predicted <ME>  
 F-390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
 F-1007-1615/Product: hepacivirus #status predicted <NS3>  
 F-1230-1237/Region: nucleotide binding motif A (P-loop)  
 F-1312-1311/Region: nucleotide binding motif B (P-loop)  
 F-1316-1319/Region: DEXH motif  
 F-1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>  
 F-1863-2013/Product: nonstructural protein NS4b #status predicted <NS4b>  
 F-196-209, 234, 250, 305, 417-423, 448, 532, 540, 556, 576, 623, 645/Binding site: carbohydrate (As  
 Query Match 93.6%; Score 1665; DB 1; Length 3010;  
 Best Local Similarity 92.7%; Pred. No. 3\_7e-132;  
 Matches 316; Conservative 10; Mismatches 15; Indels 0; Gaps 0;  
 Oy 1 VRGGDRATILLTCAVHPELIFDITKLLIAFPLMLVQOAGIKTRPVYRVAQGLIRALMV 60  
 Db 866 VRGGRDAITILLTCAVHSEFLDFITKLLIAFPLMLVQOAGILTRPVYFVRAGQILRACLMV 925  
 Qy 61 RKGAGHIVOMAKMALKALTGYVYDHITPLQWAKHGLRDAAVAPVVISDEMKHLIT 120  
 Db 926 RKVAGGHYVOMALMKALKALTGYVYDHITPLQWAKHGLRDAAVAPVVISDEMKHLIT 985  
 Qy 121 WGDATTAACGDTISGLPPVSARRGEELIGPADNPEGQWRLAPITAYSQORTGGLGCIT 180  
 Db 986 WGDATTAACGDTISGLPPVSARRGEELIGPADNPEGQWRLAPITAYSQORTGGLGCIT 1045  
 Qy 181 SLTGRDKNQVEGAAQVSTATOSFLATCVNGCWTVFHGAGSKTLAQPKGKIIQMTNTVD 240  
 Db 1046 SLTGRDKNQVEGAAQVSTATOSFLATCVNGCWTVFHGAGSKTLAQPKGKIIQMTNTVD 1105  
 Qy 241 QDILVGWQARPPGARSMTCTCGSSDLVYTRHADVPTTERRGDSRGGLSPSPVSYKGSS 300  
 Db 1106 QDLVGWQARPPGASLSTPCTYGSDDLYLVTREHADVPTTERRGDSRGGLSPSPVSYKGSS 1165  
 Qy 301 GGPLICLRCGHAVGIFRAVCTGVAKAVDFEVESMTMR 341  
 Db 1166 GGPLICLRCGHAVGIFRAVCTGVAKAVDFEVESMTMR 1205

A;Molecule type: mRNA  
A;Residues: 1-3011 <CHO>  
A;Cross-references: GB:Me62321; NID:9329873; PIDN:AAA45676.1; PID:9329874  
R;Chan, S.W.; McCormick, F.; Holmes, E.C.; Dow, B.; Peutheuer, J.F.; Follett, E.; Yap, P.I.  
J. Gen. Virol. 73, 1131-1141, 1992  
A;Title: Analysis of a new hepatitis C virus type and its phylogenetic relationship to e  
A;Reference number: PQ0393; MUID:9228871; PMID:116939  
A;Accession: PQ0403  
A;Molecule type: genomic RNA  
A;Residues: 157-1633 <CHA>  
A;Cross-references: DDBJ:010128  
A;Experimental source: isolates E-b16  
A;Accession: PQ0404  
A;Status: preliminary  
A;Molecule type: genomic RNA  
A;Residues: 157-1633 <CH2>  
A;Experimental source: isolates E-b17  
C;Superfamily: hepatitis C virus genome polyprotein  
C;Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructure  
C;Protein: capsid protein C #status predicted <CPC>  
F;115-115/Product: capsid protein C #status predicted <CPC>  
F;116-191/Product: envelope protein M #status predicted <EPM>  
F;192-389/Product: major envelope protein E #status predicted <MEB>  
F;299-729/Product: nonstructural protein NS1 #status predicted <NS1>  
F;730-1006/Product: nonstructural protein NS2 #status predicted <NS2>  
F;107-1615/Product: hepacivirin #status predicted <NS3>  
F;1230-1237/Region: nucleotide-binding motif A (P-loop)  
F;1312-1317/Region: nucleotide-binding motif B  
F;1312-1319/Region: nucleotide-binding motif B  
F;1312-1319/Region: nucleotide-binding motif B  
F;1616-1862/Product: nonstructural protein NS4a #status predicted <NS4a>  
F;1616-1862/Product: nonstructural protein NS4b #status predicted <NS4b>  
F;1863-2013/Product: nonstructural protein NS5 #status predicted <NS5>  
F;196,209,234,305,325,417,423,430,448,476,532,540,556,576,623,645,1213,1255,2041,2077,22  
Query Match 88.0%; Score 1565; DB 1; Length 301;  
Best local Similarity 84.2%; Pred. No. 1,1e-123;  
Matches 287; Conservative 28; Mismatches 26; Indels 0; Gaps 0;  
QY 1 VRGGDAATLTCAVHPPELFDITKLLAIFGPMVNLVAGITKPYVRAQIJRAMLV 60  
866 VRGGDAATLTCAVHPPELFDITKLLAIFGPMVNLVAGITKPYVRAQIJRAMLV 60  
Db 926 RKAAGHYYOMAFMVKLAALTGTYYDHPLDPLWAHQSLRDLAVAFVFSOMEKLIT 985  
Db 86 VRGGDAATLTCAVHPPELFDITKLLAIFGPMVNLVAGITKPYVRAQIJRAMLV 60  
Db 926 RKAAGHYYOMAFMVKLAALTGTYYDHPLDPLWAHQSLRDLAVAFVFSOMEKLIT 985  
Db 986 WGADTAACGDIISGLPVSAARGRREILGPADGMVSKGWRLLAPITAYAQTRGILGCIT 1045  
Db 121 WGADTAACGDIISGLPVSAARGRREILGPADGMVSKGWRLLAPITAYAQTRGILGCIT 1045  
Db 986 WGADTAACGDIISGLPVSAARGRREILGPADGMVSKGWRLLAPITAYAQTRGILGCIT 1045  
Db 181 SLTGDRKNOVEGVONSTATOSFLATATINGVQWTFVHGAGRTTASPKGPVTOMYNVD 240  
Db 1046 SLTGDRKNOVEGVONSTATOSFLATATINGVQWTFVHGAGRTTASPKGPVTOMYNVD 1105  
Db 241 QDLVQWQAPPGARSMTCTCGSSDLYLVRHADVIPRRGDSRSLSPPRPSYIKSS 300  
Db 1106 QDLVQWQAPPGARSMTCTCGSSDLYLVRHADVIPRRGDSRSLSPPRPSYIKSS 1165  
QY 301 GGPLICPSPGHAVGIFRAVCTRGVAKAVDPIPESMETMR 341  
Db 1166 GGPLICPSPGHAVGIFRAVCTRGVAKAVDPIPESMETMR 1206  
**RESULT 7**  
S40770 genome polyprotein - hepatitis C virus  
N;Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructural protein NS4a); nonstructural protein NS4b; nonstructural protein NS5  
C;Species: hepatitis C virus  
C;Date: 19-May-2000 #sequence\_revision 31-Dec-1992 #text\_change 19-Jan-2001  
C;Accession: A36814; A41546  
R;Inchoua, G.; Zelde, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.  
Submitted to GenBank, July 1992  
A;Description: Genomic structure of the human prototype strain H of hepatitis C virus: c  
A;Reference number: A36814  
A;Accession: A36814  
A;Molecule type: genomic RNA  
A;Residues: 1-3011 <INC>  
A;Reference number: S40770

A;Cross-references: GB:M67463; NID:9329737; PID:9329738  
 R;Inchaispe, G.; Zebedge, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.  
 Proc. Natl. Acad. Sci. U.S.A. 88, 10292-10296, 1991  
 A;Reference number: A41546; MUID:9205226; PMID:1658800  
 A;Contents: annotation  
 A;Note: neither amino acid nor nucleotide sequence is given  
 C;Superfamily: hepatitis C virus genome polyprotein  
 C;Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural  
 F:116-131/Product: envelope protein M #status predicted <CPC>  
 F:116-131/Product: envelope protein M #status predicted <BPM>  
 F:192-389/Product: major envelope protein E #status predicted <ME>  
 F:384-408/Region: hypervariable #status predicted  
 F:390-730/Region: nonstructural protein NS1 #status predicted <NS1>  
 F:731-1007/Product: nonstructural protein NS2 #status predicted <NS2>  
 F:1008-1616/Product: hepcivirin #status predicted <NS3>  
 F:1390-729/Product: nonstructural protein NS1 #status predicted <NS1>  
 F:730-1005/Product: nonstructural protein NS2 #status predicted <NS2>  
 F:1007-1615/Product: hepcivirin #status predicted <NS3>  
 F:1233-1238/Region: nucleotide-binding motif A (P-loop)  
 F:1311-1317/Region: nucleotide-binding motif B (P-loop)  
 F:1316-1319/Region: DEKH motif  
 F:1611-1862/Product: nonstructural protein NS4a #status predicted <NA4>  
 F:1863-2013/Product: nonstructural protein NS4b #status predicted <NA4>  
 F:1924-3011/Product: nonstructural protein NS5 #status predicted <NS5>  
 F:196,209,234,305,325,417,423,430,448,476,522,540,556,576,623,645,1213,1255,2041,2240,23  
 Query Match 87.5%; Score 1555; DB 1; Length 3011;  
 Best Local Similarity 83.9%; Pred. No. 7.6e-123; Mismatches 26; Indels 0; Gaps 0;  
 Matches 286; Conservative 29; Mismatches 26;  
 QY 1 VRGGDRATILTCVHPELIFDITKLLAIAFGPLUNVLOGITKPYFRAQGIRACMLV 60  
 Db 866 VRGERDADVLITCVHPELIFDITKLLAIFGPWLQSLSLXKPYFTRVQJRLICLA 925  
 QY 61 RKAAGHHYQAMAFMKLAATGTYVYDHLPQLQDWAHLERLDLVAVERVIFSMEVKHT 120  
 Db 926 RKAAGHHYQAMAFMKLAATGTYVYDHLPQLQDWAHLERLDLVAVERVIFSMEVKHT 985  
 QY 121 WGADTAAGGIVQAMAFMKLAATGTYVYDHLPQLQDWAHLERLDLVAVERVIFSMEVKHT 120  
 Db 987 WGADTAAGGIVQAMAFMKLAATGTYVYDHLPQLQDWAHLERLDLVAVERVIFSMEVKHT 1046  
 QY 161 SLTGRDKNQVEGVQVQVSTQSLATCNGVQWVHAGSKTLAGPKGPIQTMNTD 240  
 Db 1047 SLTGRDKNQVEGVQVLSLTQATQFLGICINGVWMTFHAGSKTLAGPKGPIQVQMTND 1105  
 QY 241 QDVGWQAPPGRASMPCTCGSSDLVTRADVPVRAGDSRSILSPRVSYLKGS 300  
 Db 1107 QDVGWQAPPGRASMPCTCGSSDLVTRADVPVRAGDSRSILSPRVSYLKGS 1166  
 QY 301 GGPLCPSPGHAVGIFRAAVCTRGVAKAVDFIPVENLETMR 341  
 Db 1166 GGPLCPSPGHAVGIFRAAVCTRGVAKAVDFIPVENLETMR 1206  
 RESULT 9  
 JCS620 genome polyprotein - hepatitis C virus (isolate EUH1480) 300  
 N;Contains: capsid protein C; envelope protein M; hepcivirin (EC 3.4.21.98) (nonstructural protein NS4a); nonstructural protein NS4b; nonstructural protein NS5  
 C;Species: hepatitis C virus  
 C;Accession: JCS620 #sequence\_revision 19-May-2000 #text\_change 19-May-2000  
 C;Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 17-Nov-2000  
 C;Accession: JQ1303  
 R;Okamoto, H.; Okada, S.; Sugiama, Y.; Kural, K.; Iizuka, H.; Machida, A.; Miyakawa, Y.;  
 J. Gen. Virol. 72, 2697-2704, 1991  
 A;Title: Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from a human  
 A;Reference number: JQ1303; MUID:9204440; PMID:1658196  
 A;Molecule type: genomic RNA  
 A;Residues: 1-3033 <OKA>  
 A;Cross-references: GB:D00944; NID:9221650; PID:9221651  
 A;Experimental source: isolate HC-J6 from a Japanese individual  
 C;Superfamily: hepatitis C virus genome polyprotein  
 C;Keywords: ATP; glycoprotein; hydrolase; P-loop; polyprotein; serine proteinase; transmembrane  
 F:2-15/Product: capsid protein C #status predicted <CPC>  
 F:116-191/Product: envelope protein M #status predicted <EM>  
 F:192-389/Product: major envelope protein E #status predicted <ME>  
 F:390-733/Product: nonstructural protein NS1 #status predicted <NS1>  
 F:734-110/Product: nonstructural protein NS2 #status predicted <NS2>  
 F:1011-1619/Product: hepcivirin #status predicted <NS3>  
 F:1311-1321/Region: nucleotide-binding motif B  
 F:1320-1323/Region: DEKH motif  
 F:1620-1865/Product: nonstructural protein NS4a #status predicted <NA4>  
 C;Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serine  
 F:2-115/Product: capsid protein C #status predicted <CPC>  
 F:1867-2017/Product: nonstructural protein NS5 #status predicted <NA5>  
 F:2210-2249/Region: interferon sensitivity determining #status predicted  
 Query Match 76.6%; Score 1362; DB 1; Length 3014;  
 Best Local Similarity 71.3%; Pred. No. 1.7e-106; Mismatches 49; Indels 0; Gaps 0;  
 Matches 243; Conservative 49; Mismatches 49; Indels 0; Gaps 0;  
 QY 1 VRGGDRATILTCVHPELIFDITKLLAIAFGPLUNVLOGITKPYFRAQGIRACMLV 60  
 Db 867 VKGRGDATVIMLCLFHPLGFEITKILFGILGPLYLQLHSITKVPIFLRALARLCLLA 926  
 QY 121 WGADTAAGGIVQAMAFMKLAATGTYVYDHLPQLQDWAHLERLDLVAVERVIFSMEVKHT 120  
 Db 927 KHLVYKVKVQALHILGRGLTGYVHAPKDWASGLRBLTVATEPTVFSAMTKTIV 986  
 QY 161 SLTGRDKNQVEGVQVQVSTQSLATCNGVQWVHAGSKTLAGPKGPIQTMNTD 240  
 Db 1047 SLTGRDKNQVEGVQVLSLTQATQFLGICINGVWMTFHAGSKTLAGPKGPIQVQMTND 1105  
 QY 241 QDVGWQAPPGRASMPCTCGSSDLVTRADVPVRAGDSRSILSPRVSYLKGS 300  
 Db 1107 QDVGWQAPPGRASMPCTCGSSDLVTRADVPVRAGDSRSILSPRVSYLKGS 1166  
 QY 301 GGPLCPSPGHAVGIFRAAVCTRGVAKAVDFIPVENLETMR 341  
 Db 1167 GGPLCPSPGHAVGIFRAAVCTRGVAKAVDFIPVENLETMR 1207  
 RESULT 10  
 JQ1303 genome polyprotein - hepatitis C virus (isolate HC-J6) 300  
 N;Contains: capsid protein C; envelope protein M; hepcivirin (EC 3.4.21.98) (nonstructural protein NS4a); nonstructural protein NS4b; nonstructural protein NS5  
 C;Species: hepatitis C virus  
 C;Accession: JQ1303 #sequence\_revision 19-May-2000 #text\_change 17-Nov-2000  
 C;Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 17-Nov-2000  
 C;Accession: JQ1303  
 R;Okamoto, H.; Okada, S.; Sugiama, Y.; Kural, K.; Iizuka, H.; Machida, A.; Miyakawa, Y.;  
 J. Gen. Virol. 72, 2697-2704, 1991  
 A;Title: Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from a human  
 A;Reference number: JQ1303; MUID:9204440; PMID:1658196  
 A;Molecule type: genomic RNA  
 A;Residues: 1-3033 <OKA>  
 A;Cross-references: GB:D00944; NID:9221650; PID:9221651  
 A;Experimental source: isolate HC-J6 from a Japanese individual  
 C;Superfamily: hepatitis C virus genome polyprotein  
 C;Keywords: ATP; glycoprotein; hydrolase; P-loop; polyprotein; serine proteinase; transmembrane  
 F:2-15/Product: capsid protein C #status predicted <CPC>  
 F:116-191/Product: envelope protein M #status predicted <EM>  
 F:192-389/Product: major envelope protein E #status predicted <ME>  
 F:390-733/Product: nonstructural protein NS1 #status predicted <NS1>  
 F:734-110/Product: nonstructural protein NS2 #status predicted <NS2>  
 F:1011-1619/Product: hepcivirin #status predicted <NS3>  
 F:1311-1321/Region: nucleotide-binding motif B  
 F:1320-1323/Region: DEKH motif  
 F:1620-1865/Product: nonstructural protein NS4a #status predicted <NA4>  
 C;Keywords: ATP; glycoprotein; hydrolase; nucleotide binding; P-loop; polyprotein; serine  
 F:2-115/Product: capsid protein C #status predicted <CPC>  
 F:1867-2017/Product: nonstructural protein NS5 #status predicted <NA5>  
 F:2210-2249/Region: interferon sensitivity determining #status predicted

F;195,209,234,305,325,417,423,430,448,477,534,542,558,578,627,649,1091,1217,1259,2038,28

Query Match 71.4%; Score 1269; DB 1; Length 3033;  
Best Local Similarity 67.2%; Pred. No. 1; 3e-98;  
Matches 229; Conservative 49; Mismatches 63; Indels 0; Gaps 0;

QY 1 VRGGDAILITCAVHPPELFDTIKLLAIFGPMVHQAGITKPYFVRAQGLRACMV 60  
Db 870 VRGGDRGIWAVATFCPGWFDITKLLAIFGPMVHQAGITKPYFVRAQGLRACMV 929  
QY 61 RKAAGHHYQOMAFMKLAALTGTYYDHJLPLQWAHAGIRDLAVAVEPVITFSMEVKIT 120  
Db 930 RHLAGGRYVOMVLALGRLGTYDHLJPPMSDWAANGIRDLAVAVEPVITFSMEVKIV 989  
QY 121 WGADTAACDIISCLPVSARRGRREILLGPADEQGQWRLAPITAVSQQRGGLCIT 180  
Db 990 WGAETAACGDIHLGPLVSPARLGREVLGPDGVTSGKWSLLAPITAVQQTGLGTTW 1049  
QY 181 SLTGRDKRNQVEGEVQVSTATOSFLATCNGVCWTVFHAGSKTLAGPKGPTQMYINVD 240  
Db 1050 SMTGDRDXTEPAGEQVLSTWTSQSLGTTGIVLWVTHAGNKLASLAPIFYRAHALLRCMV 929  
QY 241 QDLVQWQAPPGRASMPCTCGSSDLYLVTRHADYLVRERGDGSLSLSPRVSYLKGS 300  
Db 1110 GDLVQWPSPPGKSLDPCGAVDLYLVTRHADYLVRERGDGSLSLSPRVSYLKGS 1169  
QY 301 GGPLICPSPGHAVGIGRAVCTRGVAKAVDIPVPMETMWR 341  
Db 1170 GGPVILCPRGHGVFRAAVCSRGVAKSIDFIPVETLIVTR 1210

RESULT 11  
GNWJB

genome polyprotein - hepatitis C virus (strain HC-J8)  
N;Contains: capsid protein C; envelope protein M; hepacivirin (EC 3.4.21.98) (nonstructural protein NS4a); nonstructural protein NS4b; nonstructural protein NS5  
C;Species: hepatitis C virus  
C;Date: 31-Dec-1992 #sequence\_revision 31-Dec-1992 #text\_change 19-Jan-2001  
C;Accession: A40250; PQ0397; PQ0559  
R;Ozamoto, H.; Kurai, K.; Okada, S.I.; Yamamoto, K.; Lizuka, H.; Tanaka, T.; Fukuda, S.; Viriology 188, 331-341, 1992  
A;Title: Full-length sequence of a hepatitis C virus genome having poor homology to reported reference number: A40250; MUID:92230232; PMID:131459  
A;Accession: A40250  
A;Molecule type: genomic RNA  
A;Residues: 1-3033 <OKA>  
A;Cross-references: GB:D10988; GB:D01221; NID:9221608; PIDN:BAA01761.1; PID:9221609  
R;Chan, S.W.; McOmisch, F.; Holmes, E.C.; Dow, B.; Peitlheuer, J.F.; Follett, E.; Yap, P.L.; Gen. Virol. 73, 1131-1141, 1992  
A;Title: Analysis of new hepatitis C virus type and its phylogenetic relationship to existing types: comparison of the genome of hepatitis C virus type 1a with type 1b  
A;Reference number: PQ0393; MUID:9226871; PMID:1316939  
A;Accession: PQ0393  
A;Molecule type: genomic RNA  
A;Residues: 2078-2754 <CHA>  
A;Cross-references: DDB:D10134  
A;Experimental source: Isolate E-b12  
R;Koto, N.; Ootsuyama, Y.; Ohkoshi, S.; Nakazawa, T.; Mori, S.; Hijikata, M.; Shimotohno, K.; Biophys. Res. Commun. 181, 279-285, 1991  
A;Title: Distribution of pluri-HCV types in Japan.  
A;Reference number: PQ0554; MUID:92068204; PMID:1720309  
A;Accession: PQ0559  
A;Molecule type: mRNA  
A;Residues: 2078-2729 <CAT>  
A;Cross-references: GB:D10562; GB:D90518; NID:9221523; PIDN:BA01418.1; PID:9221524  
C;Superfamily: hepatitis C virus genome polyprotein  
C;Keywords: ATP; capsid protein; envelope protein; glycoprotein; hydrolase; nonstructural protein; Product: capsid protein C #status predicted <CPC>  
F;1-115/Product: envelope protein M #status predicted <EPM>  
F;116-197/Product: envelope protein E #status predicted <ME>  
F;198-389/Product: nonstructural protein NS1 #status predicted <NS1>  
F;734-1010/Product: nonstructural protein NS2 #status predicted <NS2>  
F;1011-1619/Product: hepacivirin #status predicted <NS3>  
F;1234-1241/Region: nucleotide-binding motif A (p-loop)

F;1316-1321/Region: nucleotide-binding motif B

Query Match 71.0%; Score 1262; DB 1; Length 3033;  
Best Local Similarity 66.0%; Pred. No. 4; 9e-98;  
Matches 225; Conservative 54; Mismatches 62; Indels 0; Gaps 0;

QY 1 VRGGDAILITCAVHPPELFDTIKLLAIFGPMVHQAGITKPYFVRAQGLRACMV 60  
Db 870 VRGGDRGIWAVATFCPGWFDITKLLAIFGPMVHQAGITKPYFVRAQGLRACMV 929  
QY 61 RKAAGHHYQOMAFMKLAALTGTYYDHJLPLQWAHAGIRDLAVAVEPVITFSMEVKIT 120  
Db 930 RHLAGGRYVOMVLALGRLGTYDHLJPPMSDWAANGIRDLAVAVEPVITFSMEVKIV 989  
QY 121 WGADTAACDIISCLPVSARRGRREILLGPADEQGQWRLAPITAVSQQRGGLCIT 180  
Db 990 WGAETAACGDIHLGPLVSPARLGREVLGPDGVTSGKWSLLAPITAVQQTGLGTTW 1049  
QY 181 SLTGRDKRNQVEGEVQVSTATOSFLATCNGVCWTVFHAGSKTLAGPKGPTQMYINVD 240  
Db 1050 SMTGDRDXTEPAGEQVLSTWTSQSLGTTGIVLWVTHAGNKLASLAPIFYRAHALLRCMV 929  
QY 241 QDLVQWQAPPGRASMPCTCGSSDLYLVTRHADYLVRERGDGSLSLSPRVSYLKGS 300  
Db 1110 GDLVQWPSPPGKSLDPCGAVDLYLVTRHADYLVRERGDGSLSLSPRVSYLKGS 1169  
QY 301 GGPLICPSPGHAVGIGRAVCTRGVAKAVDIPVPMETMWR 341  
Db 1170 GGPVILCPRGHGVFRAAVCSRGVAKSIDFIPVETLIVTR 1210

RESULT 12  
T08841

polyprotein - douroucouli hepatitis B virus A  
C;Species: douroucouli hepatitis B virus A  
C;Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 17-Nov-2000  
C;Accession: T08841  
R;Berker, J.C.; Desai, S.M.; Leary, T.P.; Chalmers, M.L.; Montes, C.C.; Musahawar, I.K.; Gen. Virol. 79, 41-45, 1998  
A;Title: Genomic analysis of two GB virus A variants isolated from captive monkeys.  
A;Reference number: Z16486; MUID:98120818; PMID:9460920  
A;Accession: T08841  
A;Status: translated from GB/EMBL/DDBJ  
A;Molecule type: mRNA  
A;Residues: 1-1005 <ERK>  
A;Cross-references: EMBL:AF023425; NID:9282599; PIDN: AAC40502.1; PID:92828600  
C;Superfamily: hepatitis C virus genome polyprotein  
C;Keywords: polyprotein

Query Match 23.0%; Score 408.5; DB 2; Length 3005;  
Best Local Similarity 31.3%; Pred. No. 8; 9e-26;  
Matches 111; Conservative 57; Mismatches 154; Indels 33; Gaps 9;

QY 2 RGRD---AIIILCAVHPPELFDTIKLLAIFGPMVHQAGITKPYFVRAQGLRAC 57  
Db 813 RGRDWRVTAVWAAAGIFPREWRVCSAALTA-----LAALDSIVLLET-ILTA 864

QY 58 MVRKA-----AGHYQOMAFMKLAALTGTYYDHJLPLQWAHAGIRDLAVAVEPV 109  
Db 865 QPARAARLRLSITLFGDADLTFRRRLERGRVTLFOHQGQVSXGAALXDLGVALPY 924

QY 110 IFSDMEVKITWGAATACGDIISGIPVSARRGRREILLG--PADNFEGQWRLAPITAV 167  
Db 925 SVTARQYIVEDAARTACGQSVLPVWRGRDVIVFPSTRALPQGPVTPAPVV- 983

QY 168 SOQTRGLGCHITSURGKNOVEGEVQVSTATOSFLATCNGVCWTVFHAGSKTLAG 227  
Db 984 MORGJGFPSVVKTSMLGRDERHEGSIWVIGTISTRSGTCVNGVWITPHGSNARTAG 1043

C;Keywords: structural protein  
 F;1-22;Domain: signal sequence #status predicted <SIG>  
 F;23-660;Product: structural protein 2 #status predicted <SP2>  
 Query Match 5.8%; Score 102.5; DB 1; Length 660;  
 Best Local Similarity 19.8%; Pred. No. 1.1; Gaps 19;  
 Matches 82; Conservative 52; Mismatches 130; Indels 151; Gaps 19;  
 RESULT 13  
 T0839 polyprotein - marmoset hepatitis GB virus A  
 C;Species: marmoset hepatitis GB virus A  
 C;Accession: T0839  
 Date: 23-Jul-1999 #sequence\_revision 23-Jul-1999 #text\_change 17-Nov-2000  
 R;Birker, J.C.; Debei, S.M.; Leary, T.P.; Chalmers, M.L.; Montes, C.C.; Mushahwar, I.K.  
 J; Gen. Virol. 79, 41-45, 1998  
 A;Title: Genomic analysis of two GB virus A variants isolated from captive monkeys.  
 A;Reference number: Z15486; MUID:98120818; PMID:9460920  
 A;Accession: T0839  
 A;Status: translated from GB/EMBL/DDBJ  
 A;Molecule type: genomic RNA  
 A;Residues: 1-2970 <ERK>  
 A;Cross-references: EMBL:AF023424; NID:92828597; PIDN:AC40501.1; PID:92828598  
 C;Superfamily: hepatitis C virus genome polyprotein  
 C;Keywords: polyprotein  
 Query Match 19.3%; Score 342.5; DB 2; Length 2970;  
 Best Local Similarity 28.9%; Pred. No. 3.4e-20; Gaps 12;  
 Matches 103; Conservative 56; Mismatches 133; Indels 65; Gaps 12;  
 QY 26 LLLATFGP-----IMVQAGITKVPKVFRAGGLIRACMLYKAAGCHYVONAFMKA 77  
 816 LVAAMFWPRBIAAVCAFLGFFDWDYILEVNLVNSPNLVLRARVLDLVAGDKA 875  
 Db 78 ALTGVVVVYVHPTPLQD-WAHG-----RDLAYAVEPTFSDMEVKIITWGADTA 126  
 Db 876 T---TWLVERKLKRKNCFLYAHAGQVTRTAEQLQWQGALEPVAVHPEDCAMVRDAETL 932  
 Db 877 AGCDIISGLPVSARRGREGIELGPADNFECOGWRU-----LAPITAVSQTRGELGCIT 180  
 Qy 933 SCGGSVHKGPKVVARRGDEVLEVLGVINGV----WELLPPGFVPTAPVWV-HHKGKPFVGVKT 987  
 QY 181 SLTGRDKNQVEGEQVQVSTATQSFLATCUNGCVTFVHGAGSKTLAGKGPKTQMYITVD 240  
 988 SMTGMDDETERGVNTVWLGSTRSMGTCTVNGVMTTYHGSNARTLAQMGPPNSRWNSA 1047  
 QY 241 QDLVQWQAPPGRASMTPTCGSSDIYLVTRHADVIPVRRGDSRGSLIS----- 289  
 Db 1048 DDVAVYPLVPGAKCPCQCPQGWVI-----RND-GALCHGTLGRTVBDL 1094  
 QY 290 PRPVSYLKGSSSGGPLCPSPGHAVGIFRAAVCTRG-----VARAVDIPVSENETT 339  
 Db 1095 PAAELCDPFRGSSGSPICDCBHAVGML-TSVLHRGSRVTGIRYTKPWEILPREAITHT 1150  
 RESULT 14  
 VHHWH2  
 structural protein 2 precursor - hepatitis E virus (strain Burma)  
 C;Species: hepatitis E virus  
 C;Date: 30-Sep-1992 #sequence\_revision 30-Sep-1992 #text\_change 23-Jul-1999  
 C;Accession: C40778  
 R;Tam, A.W.; Smith, M.M.; Guerra, M.E.; Huang, C.C.; Bradley, D.W.; Fry, K.R.; Reyes, G.  
 Virology 185, 120-131, 1991  
 A;Title: Hepatitis E virus (HEV): molecular cloning and sequencing of the full-length vi  
 A;Reference number: A40778; MUID:92024067; PMID:1926770  
 A;Accession: C40778  
 A;Molecule type: genomic RNA  
 A;Residues: 1-660 <TAM>  
 A;Cross-references: GB:M73218; NID:9330023; PIDN:AA457351; PID:9330026  
 A;Note: the authors translated the codon CGC for residue 2 as Ala  
 C;Superfamily: hepatitis E virus structural protein 2  
 Query Match 5.7%; Score 101; DB 2; Length 564;  
 Best Local Similarity 26.1%; Pred. No. 1.2; Gaps 11;  
 Matches 54; Conservative 37; Mismatches 74; Indels 42; Gaps 11;

QY 15 VHPERI----FDITKULLAIFGPLMVLONGI-----TKUPYFVRAQGLIRACMLVR 61  
 Db 154 VHDDELVKLKGEGEVSBLQFAKSPTVILAGLQGVGKTTVCAKLCYLKRG--KSCMLI- 210  
 QY 62 KAAGCHYVOMAFMKLAAL--TGIVYVYDHITPLQ--DWAHAGIRDЛАVАVEPVIFSDMEV 116  
 Db 211 --AGDVYRPAIDOLVILGEQVGPVYITDVKPADIAKQGLKEAKK---NNVDV 261  
 QY 117 KIRIWGADAAACGDIISGFLVSARGRELL----LGPADENFECOGWRLLAPITAYQQT 171  
 Db 262 VIM----DIRGLQIDKGMDELKDKVKEFLNPTEVLLVDAAMTCQ - BAAALVTFNVEI 315  
 QY 172 RGLIGCIRISLTGDKDKNYEGEYVWVS 198  
 Db 316 -GIGAILTRLDGSRGGAALSVKEVS 341

Search completed: May 6, 2004, 09:37:20  
 Job time : 12:1056 secs

Copyright (c) 1993 - 2004 Compugen Ltd.

### CM protein - protein search, using sw model

Run on: May 6, 2004, 09:09:55 ; Search time 7.119 Seconds

Perfect score: US-10-650-585-14

Sequence: 1 VRGGDAAILLTCATHPELT.....RGVAKAVDIFPVESMETMNR 341

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched:

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
listing first 45 summaries

Database : Swissprot\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

### SUMMARIES

| Result No. | Score | Query Match Length | DB ID | Description   |
|------------|-------|--------------------|-------|---------------|
| 1          | 1711  | 96.2               | 3010  | 1 P0LG_HCVJT  |
| 2          | 1710  | 96.2               | 3010  | 1 P0LG_HCVJA  |
| 3          | 1685  | 94.8               | 3010  | 1 P0LG_HCVTW  |
| 4          | 1649  | 92.7               | 3010  | 1 P0LG_HCVBK  |
| 5          | 1565  | 88.0               | 3011  | 1 P0LG_HCVI   |
| 6          | 1555  | 87.5               | 3011  | 1 P0LG_HCVH   |
| 7          | 1269  | 71.4               | 3033  | 1 P0LG_HCVJS  |
| 8          | 1262  | 71.0               | 3033  | 1 P0LG_HCVB   |
| 9          | 102.5 | 5.8                | 660   | 1 VST2_HEVPA  |
| 10         | 102.5 | 5.8                | 660   | 1 VST2_HEVPA  |
| 11         | 101   | 5.7                | 564   | 1 SR5C_ARATH  |
| 12         | 101   | 5.7                | 600   | 1 DPO2_MOUSE  |
| 13         | 93.5  | 5.3                | 1380  | 1 CYLA_LIENDO |
| 14         | 93    | 5.2                | 434   | 1 TCOL_CHLTO  |
| 15         | 92.5  | 5.2                | 706   | 1 TRFE_HORSE  |
| 16         | 92    | 5.2                | 659   | 1 VST2_HEVME  |
| 17         | 91.5  | 5.1                | 485   | 1 VST2_HEVRHE |
| 18         | 91.5  | 5.1                | 660   | 1 VST2_HEVMY  |
| 19         | 91.5  | 5.1                | 3414  | 1 P0LG_TBENVW |
| 20         | 90.5  | 5.1                | 3414  | 1 P0LG_TBENV  |
| 21         | 89.5  | 5.0                | 961   | 1 ATCT_YERPE  |
| 22         | 87.5  | 4.9                | 3412  | 1 P0LG_TBENV  |
| 23         | 86.5  | 4.9                | 347   | 1 MDHM_EUCGU  |
| 24         | 86.5  | 4.9                | 470   | 1 NRAM_IAWHM  |
| 25         | 86    | 4.8                | 338   | 1 GALE_NEIGO  |
| 26         | 86    | 4.8                | 433   | 1 DCUA_WOLSU  |
| 27         | 85    | 4.8                | 470   | 1 NRAM_TATRA  |
| 28         | 85    | 4.8                | 730   | 1 HELS_METRA  |
| 29         | 85    | 4.8                | 854   | 1 PWP2_SCHPO  |
| 30         | 85    | 4.8                | 3313  | 1 CIR3_RAT    |
| 31         | 84.5  | 4.8                | 1705  | 1 PTPI_MOUSE  |
| 32         | 84    | 4.7                | 309   | 1 UCP2_RAT    |
| 33         | 84    | 4.7                | 1     | GPDA_COREF    |

### ALIGNMENTS

|          |   |       |            |                         |      |          |
|----------|---|-------|------------|-------------------------|------|----------|
| RESULT 1 | P0LG_HCVJT  | ID    | P0LG_HCVJT | STANDARD;               | PRT; | 3010 AA. |
| AC       | Q00269  | h     | genome     | po                      |      |          |
| DT       | 01-APR-1993   | (Rel. | 25,        | Created)                |      |          |
| DT       | 01-APR-1993   | (Rel. | 25,        | Last sequence update)   |      |          |
| DR       | 10-OCT-2003   | (Rel. | 42,        | Last annotation update) |      |          |
| DE       | Genome polypeptide C (contains: Capsid protein C (Core protein) (P22); Envelope glycoprotein E1 (GP32); Envelope glycoprotein E2 (GP68); GP70) (NS1); Protein p7; Nonstructural protein NS2 (P21); DE (EC 3.4.22.-); Protease/helicase protein NS3 (P70) (Hepacivirus)  |       |            |                         |      |          |
| DE       | (EC 3.4.21.-); Nonstructural Protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P70) (RNA-directed RNA Polymerase) (EC 2.7.7.48)).  |       |            |                         |      |          |
| OS       | Hepatitis C virus (isolate HC-JT) (HCV).  |       |            |                         |      |          |
| OC       | Viruses; sRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.   |       |            |                         |      |          |
| OU       | OK  |       |            |                         |      |          |
| RN       | NCBI_TaxID:31642;   |       |            |                         |      |          |
| [1]      |   |       |            |                         |      |          |
| RP       | SEQUENCE FROM N.A.  |       |            |                         |      |          |
| RX       | MEDLINE=92295714; PubMed=1318627; R   |       |            |                         |      |          |
| RA       | Tanaka T., Kato M., Nakagawa M., Octaviana Y., Cho M.J., Nakazawa T., Hijikata M., Ishimura Y., Shimokohno K.;  |       |            |                         |      |          |
| RT       | "Molecular cloning of hepatitis C virus genome from a single Japanese carrier: sequence variation within the same individual and among infected individuals";   |       |            |                         |      |          |
| RU       | Virus Res. 23:39-53(1992).  |       |            |                         |      |          |
| CC       | -I- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are hydrophobic, suggesting a possible membrane-related function. NS3 and NS5 may play a role in the viral RNA replication.  |       |            |                         |      |          |
| CC       | -I- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polyprotein, commonly with Asp or Glu in the P6 position, Cys or Thr in P1 and Ser or Ile in P1'.   |       |            |                         |      |          |
| CC       | -I- CATALYTIC ACTIVITY: Nucleoside triphosphate = N diphosphate + (RNA) (N).  |       |            |                         |      |          |
| CC       | -I- SUBUNIT: The virion of this virus is a nucleocapsid covered by a lipoprotein envelope. The envelope consists of two proteins: protein M and glycoprotein E. The nucleocapsid is a complex of protein C and mRNAs.   |       |            |                         |      |          |
| CC       | -I- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.   |       |            |                         |      |          |
| CC       | This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to license@isb-sib.ch). |       |            |                         |      |          |
| CC       | EMBL: D11168; BAA01943.1; -.  |       |            |                         |      |          |
| DR       | PRTR: A45573; A45573.   |       |            |                         |      |          |
| DR       | MEROPS: S29_001; -.   |       |            |                         |      |          |
| DR       | IPR000003; CYS_Ser_trypsin.   |       |            |                         |      |          |
| DR       | IPR001410; DEAD.  |       |            |                         |      |          |
| DR       | InterPro; IPR002522; HCV_capsid.  |       |            |                         |      |          |



use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; D90208; BA14233.1; -.

DR PIR; A39253; GNWV CJ.

DR HSSP; P26663; IUPA.

DR MRROPS; U39\_001; -.

DR InterPro; IPR00903; Cys\_Ser\_trypsin.

DR InterPro; IPR001410; DEAD.

DR InterPro; IPR00522; HCV\_capsid.

DR InterPro; IPR00221; HCV\_core.

DR InterPro; IPR00219; HCV\_envelope.

DR InterPro; IPR002531; HCV\_NS1.

DR InterPro; IPR002518; HCV\_NS2.

DR InterPro; IPR000745; HCV\_NS4a.

DR InterPro; IPR001490; HCV\_NS4b.

DR InterPro; IPR002868; HCV\_NS5a.

DR InterPro; IPR002166; HCV\_RNAP.

DR InterPro; IPR001650; Helicase\_C.

DR InterPro; IPR004109; Peptidase\_C29.

DR InterPro; IPR007095; RNA\_pol\_Ds\_Ps.

DR InterPro; IPR0007094; RNA\_poli\_PeVir.

DR PF01543; HCV\_capsid; 1.

DR PF01512; HCV\_core; 1.

DR PF01339; HCV\_env; 1.

DR PF01560; HCV\_NS1; 1.

DR PF01538; HCV\_NS2; 1.

DR PF02907; HCV\_NS3; 1.

DR PF01006; HCV\_NS4a; 1.

DR PF01001; HCV\_NS4b; 1.

DR PF01505; HCV\_NS5a; 1.

DR PF00271; helicase\_C; 1.

DR PRODOM; PD186062; HCV\_NS1; 1.

KW Polyprotein: Glycoprotein; Transferase; RNA-directed RNA polymerase; Core protein; Coat protein; Envelope protein; Helicase; ATP-binding; Transmembrane; Nonstructural protein; Hydrolase; Serine protease.

FT INIT\_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE CELLULAR AMINOPEPTIDASE.

FT CHAIN 1 15 CASPID PROTEIN C (POTENTIAL).

FT CHAIN 116 191 MATRIX PROTEIN (POTENTIAL).

FT CHAIN 383 729 MAJOR ENVELOPE PROTEIN E (POTENTIAL).

FT CHAIN 730 1006 NON-STRUCTURAL PROTEIN NS1 (POTENTIAL).

FT CHAIN 1007 1616 PROTEASE/HELICASE NS3 (POTENTIAL).

FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).

FT CHAIN 347 369 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).

FT TRANSMEM 1083 1083 ACT-SITE (BY SIMILARITY).

FT ACT-SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT-SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT NP-BIND 1230 1237 ACTP (POTENTIAL).

FT SITE 1316 1319 DECH BOX.

FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 234 234 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 305 305 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 423 423 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 448 448 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 532 556 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 576 623 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 623 645 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 2041 2041 N-LINKED (GLCNAC. . .) (POTENTIAL).

RESULT 3

POLG\_HCVTW ID POLG\_HCVTW STANDARD; PRT; 3010 AA.

AC P29846; DT 01-APR-1993 (Rel. 25, Last sequence update)

DT 01-APR-1993 (Rel. 25, Last annotation update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Genome polyprotein [Contains: Capsid protein C (Core Protein) (P22); Envelope glycoprotein [Contains: Capsid protein E2 (GP35); Envelope glycoprotein El (GP32); Envelope glycoprotein NS1 (GP10); Protein P7; Nonstructural protein NS2 (P21); DE (GP68); (GP70)] (NS1); Protein P7; Nonstructural protein NS2 (P21); DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirus) (EC 3.4.21.-); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA Polymerase) (EC 2.7.7.48)].

DE NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA Polymerase) (EC 2.7.7.48)].

OS Hepatitis C virus (isolate Taiwan) (HCV).

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepaciviridae; Hepacivirus.

OC NCBI TaxID=31645;

RN [1]

RR SEQUENCE FROM N.A.

RX MEDLINE#92230206; PubMed=1314449;

RA Chen P.J., Lin M.H., Tai K.F., Liu P.C., Lin C.J., Chen D.S.; RT "The Taiwanese hepatitis C virus genome: sequence determination and mapping the 5' termini of viral genomic and anti-genomic RNA.";

RL Virology 188:107-111(1992).

CC -!- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are hydrophobic, suggesting a possible membrane-related function. NS3 and NS5 may play a role in the viral RNA replication.

CC -!- CATALYTIC ACTIVITY: Nucleotide triphosphate + {RNA} (N).

CC -!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a lipoprotein envelope. The envelope consists of two proteins: protein M and glycoprotein E. The nucleocapsid is a complex of protein C and mRNA.

CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

---

EMBL; MB0754; - ; NOT\_ANNOTATED\_CDS.  
PIR; A40444; GNWV1.  
DR; IN6; 22-FEB-03.  
DR; IN5; 00-APR-98.  
MEROPS; S29\_001; -.  
MEROPS; U39\_001; -.  
DR; InterPro; IPRO09003; Cys Ser\_trypsin.  
DR; InterPro; IPRO0110; DEBD.  
DR; InterPro; IPRO0222; HCV\_capsid.  
DR; InterPro; IPRO0231; HCV\_core.  
DR; InterPro; IPRO02319; HCV\_env.  
DR; InterPro; IPRO02531; HCV\_NS1.  
DR; InterPro; IPRO02518; HCV\_NS2.  
DR; InterPro; IPRO00745; HCV\_NS4a.  
DR; InterPro; IPRO01490; HCV\_NS5b.  
DR; InterPro; IPRO02868; HCV\_NS5a.  
DR; InterPro; IPRO02166; HCV\_Rnp.  
DR; InterPro; IPRO01650; Helicase\_C.  
DR; InterPro; IPRO04109; Peptidase\_C29.  
DR; InterPro; IPRO07095; RNA\_Polymerase\_PS.  
DR; InterPro; IPRO07034; RNA\_Polymerase\_PStir.  
PFam; PRO1543; HCV\_capsid; 1.  
DR; Pfam; PRO1539; HCV\_env; 1.  
DR; Pfam; PRO1560; HCV\_NS1; 1.  
DR; Pfam; PRO1538; HCV\_NS2; 1.  
DR; Pfam; PRO02907; HCV\_NS3; 1.  
DR; Pfam; PRO1006; HCV\_NS4a; 1.  
DR; Pfam; PRO1506; HCV\_NS5a; 1.  
DR; Pfam; PRO0271; Helicase\_C; 1.  
DR; Pfam; PRO0998; Viral\_Rnp; 1.  
DR; SMART; SM00487; DEBPC; 1.  
KW; Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;  
KW; Core\_protein; Envelope\_protein; Helicase; ATP-binding;  
KW; Transmembrane; Nonstructural\_protein; Hydrolase; Serine\_protease;  
KW; 3D-structure.

INIT\_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE  
CELLULAR AMINOPEPTIDASE.

FT CHAIN 1 115 CORE PROTEIN (POTENTIAL).  
FT CHAIN 116 191 MATRIX PROTEIN (POTENTIAL).  
FT CHAIN 192 383 MATRIX ENVELOPE PROTEIN E (POTENTIAL).  
FT CHAIN 384 729 NONSTRUCTURAL PROTEIN NS1\_E2 (POTENTIAL).  
FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).  
FT CHAIN 1007 1615 PROTEASE/HELICASE NS3 (POTENTIAL).  
FT CHAIN 1616 1862 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).  
FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).  
FT CHAIN 2014 347 RNA-DIRECTED RNA POLYMERASE (POTENTIAL).  
FT TRANSMEM 1083 1083 POTENTIAL.  
FT ACT\_SITE 1107 1107 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT ACT\_SITE 1165 1165 CHARGE RELAY SYSTEM (BY SIMILARITY).  
FT NP\_BIND 1230 1237 TP (POTENTIAL).  
FT SITE 1316 1319 DECH\_BK.  
FT CARBOHYD 196 196 N-LINKED (GLCNAC . .) (POTENTIAL).  
FT CARBOHYD 209 209 N-LINKED (GLCNAC . .) (POTENTIAL).  
FT CARBOHYD 233 233 N-LINKED (GLCNAC . .) (POTENTIAL).  
FT CARBOHYD 234 250 N-LINKED (GLCNAC . .) (POTENTIAL).  
FT CARBOHYD 305 305 N-LINKED (GLCNAC . .) (POTENTIAL).  
FT CARBOHYD 417 417 N-LINKED (GLCNAC . .) (POTENTIAL).  
FT CARBOHYD 423 423 N-LINKED (GLCNAC . .) (POTENTIAL).  
FT CARBOHYD 430 430 N-LINKED (GLCNAC . .) (POTENTIAL).  
FT CARBOHYD 448 448 N-LINKED (GLCNAC . .) (POTENTIAL).

RESULT 4  
POLG\_HCVBK  
ID\_POLG\_HCVBK STANDARD; PRT; 3010 AA.  
AC\_P26673; DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Genome polyprotein [Contains: Capsid protein C (Core protein) (P22); Envelope glycoprotein E1 (GP32); Envelope glycoprotein E2 (GP48); (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21) (EC 3.4.-22.-); Protease/helicase NS3 (P70) (Hepacivirus); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)].  
DB NS4B (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48).  
OS Hepatitis C virus (isolate BK) (HCV).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus; OX NCBI\_TaxID:1105; RN [1]  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91140698; PubMed=1917440;  
RA Takemizawa A., Mori C., Fukie I., Manabe S., Murakami S., Fujita J., Onishi E., Ando T., Yoshida I., Okamura H.; "Structure and organization of the hepatitis C virus genome isolated from human carriers." J. Virol. 65:1105-1113 (1991).  
RN [2]  
RP SEQUENCE OF 1487-1500.  
RX MEDLINE=916231224; PubMed=8647104;  
RA Borowski P., Heiland M., Oehlmann K., Becker B., Kurneteky L.; "Non-structural protein 3 of hepatitis C virus inhibits

RT phosphorylation mediated by cAMP-dependent protein kinase.";  
 RN Eur. J. Biochem. 237:611-618(1996).  
 RP [3]  
 X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 1027-1215.  
 RX MEDLINE:97015088; PubMed=8161916;  
 RA Love R.A., Parge H.E., Wicksell J.A., Hostomsky Z., Habuka N.,  
 RA Moonaw E.W., Adachi T., Hostomská Z.;  
 RT "The crystal structure of hepatitis C virus NS3 protease reveals a  
 trypsin-like fold and a structural zinc binding site.",  
 RL Cell 87:331-342(1996).  
 RN [4]  
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1027-1210 AND 1678-1691.  
 RX MEDLINE:98227846; PubMed=9568917;  
 RA Yan Y., Li Y., Munsu S., Sardana V., Cole J.L., Sardana M.,  
 RA Steinkuhler C., Tomei L., de Francesco R., Kuo L.C., Chen Z.,  
 RT "Complex of NS3 protease and NS4A peptide of BK strain hepatitis C  
 virus: a 2.2-A resolution structure in a hexagonal crystal form.",  
 RL Protein Sci. 7:837-847(1998).  
 CC -!- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are  
 hydrophobic, suggesting a possible membrane-related function. NS3  
 CC and NS5 may play a role in the viral RNA replication.  
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
 precursor polyprotein, commonly with Asp or Glu in the P6  
 position, Cys or Thr in P1 or Ser or Ala in P1'.  
 CC -!- CATALYTIC ACTIVITY: Nucleoside triphosphate = N diphosphate +  
 {RNA} (N).  
 CC -!- SUBSTRATE: The virion of this virus is a nucleocapsid covered by a  
 lipoprotein envelope. The envelope consists of two proteins:  
 CC protein M and Glycoprotein E. The nucleocapsid is a complex of  
 protein C and mRNAs.  
 CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to license@isb-sib.ch).  
 DR EMBL; M58335; AAA72945.1; -.  
 DR PIR; A38455; GNWTC.  
 DR PDB; 1A1Q; 25-MAR-98.  
 DR PDB; 1JXP; 14-JAN-98.  
 DR PDB; 1MS3; 08-APR-98.  
 DR PDB; 1CPF; 15-NOV-00.  
 DR PDB; 1G3J; 08-NOV-99.  
 DR PDB; 1GX5; 09-APR-02.  
 DR DBI; 1K66; 10-APR-02.  
 DR PDB; 1QVY; 26-JUN-00.  
 DR PDB; 80RM; 20-APR-99.  
 DR MEOPS; S29.001; -.  
 DR MEROPS; U39.001; -.  
 DR InterPro; IPR009003; Cys Ser\_trypsin.  
 DR InterPro; IPR001410; DRA.D.  
 DR InterPro; IPR002522; HCV\_capsid.  
 DR InterPro; IPR002521; HCV\_core.  
 DR InterPro; IPR002519; HCV\_env.  
 DR InterPro; IPR002511; HCV\_NS1.  
 DR InterPro; IPR002518; HCV\_NS2.  
 DR InterPro; IPR000745; HCV\_NS3a.  
 DR InterPro; IPR001490; HCV\_NS3b.  
 DR InterPro; IPR002888; HCV\_NSA.  
 DR InterPro; IPR002166; HCV\_RNP.  
 DR InterPro; IPR004109; Peptidase\_C29.  
 DR InterPro; IPR007055; RNA\_Pol\_DS\_PS.  
 DR InterPro; IPR007054; RNA\_Pol\_PSVir.  
 DR Pfam; PF01543; HCV\_capsid; 1.  
 DR Pfam; PF01522; HCV\_core; 1.  
 DR Pfam; PF01539; HCV\_env; 1.  
 DR Pfam; PF01560; HCV\_NS1; 1.  
 DR Pfam; PF01538; HCV\_NS2; 1.  
 DR Pfam; PF02907; HCV\_NS3; 1.

---

DR Pfam; PF01006; HCV\_NS4a; 1.  
 DR Pfam; PF01001; HCV\_NS4b; 1.  
 DR Pfam; PF01506; HCV\_NS5a; 1.  
 DR Pfam; PF00991; Viral\_RBP; 1.  
 DR Prodrom; P0186062; HCV\_NS1; 1.  
 DR SMART; SM00487; DEDCc; 1.  
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;  
 Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;  
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease;  
 KW 3D-structure; 1  
 FT INIT\_MER; 1  
 FT FT  
 PT CHAIN 1 115  
 PT CHAIN 116 191  
 PT CHAIN 192 383  
 PT CHAIN 384 729  
 PT CHAIN 730 1006  
 PT CHAIN 1007 1615  
 PT CHAIN 1616 1862  
 PT CHAIN 1863 2013  
 PT CHAIN 2014 3010  
 PT CHAIN 3010 3369  
 PT TRANSMEM 347  
 PT ACT\_SITE 1083 1083  
 PT ACT\_SITE 1107 1107  
 PT ACT\_SITE 1165 1165  
 PT NP\_BIND 1230 1237  
 PT SITE 1316 1319  
 PT CARBOHYD 196 196  
 PT CARBOHYD 209 209  
 PT CARBOHYD 234 234  
 PT CARBOHYD 250 250  
 PT CARBOHYD 305 305  
 PT CARBOHYD 417 417  
 PT CARBOHYD 423 423  
 PT CARBOHYD 430 430  
 PT CARBOHYD 448 448  
 PT CARBOHYD 532 532  
 PT CARBOHYD 540 540  
 PT CARBOHYD 556 556  
 PT CARBOHYD 576 576  
 PT CARBOHYD 623 623  
 PT CARBOHYD 645 645  
 PT CARBOHYD 2041 2041  
 PT CARBOHYD 2077 2077  
 PT CARBOHYD 2240 2240  
 PT CARBOHYD 2529 2529  
 PT CARBOHYD 2788 2788  
 PT STRAND 1031 1035  
 PT HELIX 1039 1047  
 PT STRAND 1050 1050  
 PT STRAND 1059 1063  
 PT STRAND 1068 1074  
 PT TURN 1075 1076  
 PT STRAND 1077 1081  
 PT HELIX 1082 1085  
 PT TURN 1086 1087  
 PT STRAND 1090 1092  
 PT STRAND 1093 1094  
 PT TURN 1095 1097  
 PT STRAND 1101 1103  
 PT HELIX 1104 1107  
 PT STRAND 1108 1112  
 PT STRAND 1120 1120  
 PT STRAND 1122 1122  
 PT STRAND 1129 1133  
 PT TURN 1135 1136  
 PT STRAND 1139 1144  
 PT STRAND 1149 1157  
 PT HELIX 1158 1161  
 PT TURN 1162 1163  
 PT TURN 1165 1166  
 PT STRAND 1168 1171  
 PT TURN 1172 1174

|                  |  |           |   |   |  |
|------------------|--|-----------|---|---|--|
| FT               | STRAND   | 1175      | 1186                                      | CC  | protein M and glycoprotein E. The nucleocapsid is a complex of   |
| FT               | TURN   | 1187      | 1188                                      | CC  | protein C and mRNA.  |
| FT               | STRAND   | 1189      | 1197                                      | CC  | -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.  |
| FT               | HELIX  | 1198      | 1202                                      | CC  | This SWISS-PROT entry is copyright. It is produced through a collaboration   |
| FT               | TURN   | 1203      | 1204                                      | CC  | between the Swiss Institute of Bioinformatics and the EMBL outstation -  |
| FT               | STRAND   | 1680      | 1688                                      | CC  | the European Bioinformatics Institute. There are no restrictions on its  |
| FT               | SEQUENCE   | 3010      | AA:                                       | CC  | use by non-profit institutions as long as its content is in no way   |
| Qy               |  | 327189    | MW:                                       | CC  | modified and this statement is not removed. Usage by and for commercial  |
|                  | Best Local Similarity  | 92.1%     | Score                                     | CC  | entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> ) |
|                  | Matches  | 313;      | Pred.                                     | CC  | or send an email to license@isb-sib.ch).   |
|                  | Conservative   | 12;       | No.                                       | CC  |  |
|                  | Mismatches   | 15;       | 2.5e-13;                                  | CC  |  |
|                  | Indels   | 0;        | Gaps                                      | CC  |  |
| Db               | 2 RGGDAILITCAHPELIFDITKULLAIFGPMVLMQLQGITKPYFRAQLIRACMLR   | 61        | PDB:                                      | EMBL; M6321; AAC5676.1;                   |  |
| Qy               | 62 KAAGGHYVQMAFMKLAALTGTIVKDHLPQLDWAHAGIRDALAVEPVIFPSMVEKITW   | 121       | PIR; J39166; GIWNC3.                      |   |  |
| Db               | 927 KVAGGHYVQMAFMKLGALTGTYVNHLPLRDLPRAGRDLAVEPVIFPSMVEKITW   | 986       | PDB; 1A1V; 16-FEB-99.                     |   |  |
| Qy               | 122 GADTAACGDTISGLPVSARRGRELEILGAGDNFEGQWLLAPITAYQQTRGSLGCITS  | 181       | PDB: 1HEI; 25-NOV-98.                     |   |  |
| Db               | 987 GADTAACGDTISGLPVSARRGRELEILGAGDNFEGQWLLAPITAYQQTRGSLGCITS  | 1046      | MERCS; S29.001; -.                        |   |  |
| Qy               | 182 LTGRDKRNQEVGEVQVSTATOSFLATCNGCWTVFEGAGSKTLAGPKGPITOMYTVQD  | 241       | DR; U39.001; -.                           |   |  |
| Db               | 1047 LTGRDKRNQEVGEVQVSTATOSFLATCNGCWTVFEGAGSKTLAGPKGPITOMYTVQD   | 1106      | DR; InterPro; IPR009003; CYS_Ser_trypsin. |   |  |
| Qy               | 242 DLVGWQAPPGARSMPCTCGSSDLYVTRHADVIPVERRGDSRGSLSPRPSYLSKGSS   | 301       | DR; InterPro; IPR001410; DEAD.            |   |  |
| Db               | 1107 DLVGWQPKPPGARSJLPCGSSDLYLVTRHADVIPVRRRGDSRGSLSPRPSYLSKGSS   | 1166      | DR; InterPro; IPR002522; HCV_capsid.      |   |  |
| Qy               | 302 GLUPCPGSHAVGTRAVYOTRGVAKAVDPFPVSMETTM  | 341       | DR; InterPro; IPR002521; HCV_core.        |   |  |
| Db               | 1167 GPLLCPPFGHVGIFRAAVCTRGVAKAVDPFPVSMETTM  | 1206      | DR; InterPro; IPR002519; HCV_env.         |   |  |
| <b>RESULT 5</b>  |  |           |   |   |  |
| <b>POLG_HCV1</b> |  |           |   |   |  |
| ID               | POLG_HCV1  | STANDARD: | PRT:                                      | 3011 AA.                                  |  |
| AC               | P26644;  |           |   |   |  |
| DT               | 01-AUG-1992 (Rel. 23, Created)   |           |   |   |  |
| DT               | 01-AUG-1992 (Rel. 23, Last Sequence update)  |           |   |   |  |
| DT               | 10-OCT-2003 (Rel. 42, Last annotation update)  |           |   |   |  |
| DE               | Genome Polyprotein [Contains: Capsid protein C (core protein) (P22); Envelope glycoprotein E1 (G32); (GP35); Envelope glycoprotein E2 (GP88) (NS1); Protein P7; Nonstructural protein NS2 (P21) (EC 3.4.22.-); Protease/helicase NS3 (P70) (Repacivirin) (EC 3.4.21.98); Nonstructural protein NS4A (PA); Nonstructural protein NS4B (P27); Nonstructural protein NSSA (P56); Nonstructural protein NSSB (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)]. |           |   |   |  |
| DE               | Hepatitis C virus (isolate 1) (HCV). Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; OC Herpesvirus.   |           |   |   |  |
| OC               | NCBI_TaxID=11104;  |           |   |   |  |
| RN               | SEQUENCE FROM N.A.   |           |   |   |  |
| RP               | MEDLINE:9112826; PubMed=1948704;   |           |   |   |  |
| RX               | [1]  |           |   |   |  |
|                  | PROTEIN SEQUENCE FROM N.A.   |           |   |   |  |
|                  | RA Choo O.-L., Richman K.H., Han J.H., Berger K., Lee C., Dong C., Gallegos C., Coit D., Medina-Selby A., Barr P.J., Weiner A.J., Bradley D.W., Kuo G., Houghton M.; "Genetic organization and diversity of the hepatitis C virus"; Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).  |           |   |   |  |
|                  | RL PROC. NATL. ACAD. SCI. U.S.A. 88:2451-2455(1991).   |           |   |   |  |
|                  | CC 1-- FUNCTION: The small proteins NS2A, NS4A and NS4B are hydrophobic, suggesting a possible membrane-related function. NS3 and NS5 may play a role in the viral RNA replication.  |           |   |   |  |
|                  | CC --!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polyprotein, commonly with Asp or Glu in the P6 position. Cys or Thr in P1 and Ser or Ala in P1'. CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + (RNA) (N).   |           |   |   |  |
|                  | CC --!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a lipoprotein envelope. The envelope consists of two proteins:  |           |   |   |  |
| <b>POLG_NSA1</b> |  |           |   |   |  |
| ID               | P01538;  |           |   |   |  |
| AC               | Pfam; PF01538; HCV_NS1; 1.   |           |   |   |  |
| DT               | Pfam; PF01539; HCV_env; 1.   |           |   |   |  |
| DR               | Pfam; PF01540; HCV_NS2; 1.   |           |   |   |  |
| DR               | Pfam; PF01541; HCV_NS4b; 1.  |           |   |   |  |
| DR               | Pfam; PF01542; HCV_core; 1.  |           |   |   |  |
| DR               | Pfam; PF01543; HCV_capsid; 1.  |           |   |   |  |
| DR               | Pfam; PF01544; HCV_ns3; 1.   |           |   |   |  |
| DR               | Pfam; PF01545; HCV_ns5a; 1.  |           |   |   |  |
| DR               | Pfam; PF01546; HCV_ns5b; 1.  |           |   |   |  |
| DR               | Pfam; PF00271; helicase_C; 1.  |           |   |   |  |
| DR               | Pfam; PF00998; viral_RdRp; 1.  |           |   |   |  |
| DR               | ProDom; PD186062; HCV_NS1; 1.  |           |   |   |  |
| DR               | SMART; SM00437; DEVD_C-1.  |           |   |   |  |
| DR               | Polyprotein; Glycoprotein; Transfase; RNA-directed RNA polymerase; Core protein; Coat protein; Envelope protein; Helicase; ATP-binding; KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease; 3D-structure.   |           |   |   |  |
| FT               | INIT_MET   | 1         | 1   |   |  |
| FT               | CHAIN  | 1         | 115                                       |   | REMOVED FROM CAPSID PROTEIN C BY THE   |
| FT               | CHAIN  | 116       | 191                                       | CAPSID PROTEIN C (POTENTIAL).             |  |
| FT               | CHAIN  | 192       | 383                                       | MATRIX PROTEIN (POTENTIAL).               |  |
| FT               | CHAIN  | 384       | 729                                       | MAJOR ENVELOPE PROTEIN E (POTENTIAL).     |  |
| FT               | CHAIN  | 730       | 1006                                      | NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL). |  |
| FT               | CHAIN  | 1007      | 1615                                      | NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).    |  |
| FT               | CHAIN  | 1616      | 1862                                      | PROTEASE/HELICASE NS3 (POTENTIAL).        |  |
| FT               | CHAIN  | 1863      | 2013                                      | NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).   |  |
| FT               | CHAIN  | 2014      | 3011                                      | NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).   |  |
| FT               | TRANSMEM   | 347       | 369                                       | RNA-DIRECTED RNA POLYMERASE (POTENTIAL).  |  |
| FT               | ACT SITE   | 1083      | 1083                                      | POTENTIAL.                                |  |
| FT               | ACT SITE   | 1107      | 1107                                      | CHARGE RELAY SYSTEM (BY SIMILARITY).      |  |
| FT               | ACT SITE   | 1165      | 1165                                      | CHARGE RELAY SYSTEM (BY SIMILARITY).      |  |
| FT               | ACT SITE   | 1230      | 1237                                      | ATP (POTENTIAL).                          |  |
| FT               | NP BIND  | 1316      | 1319                                      | DECH_BOX.                                 |  |
| FT               | SITE   | 1316      | 1319                                      | N-LINKED (GLCNAC; . .) (POTENTIAL).       |  |
| FT               | CARBONYD   | 196       | 196                                       | N-LINKED (GLCNAC; . .) (POTENTIAL).       |  |
| FT               | CARBONYD   | 209       | 209                                       | N-LINKED (GLCNAC; . .) (POTENTIAL).       |  |
| FT               | CARBONYD   | 234       | 234                                       | N-LINKED (GLCNAC; . .) (POTENTIAL).       |  |
| FT               | CARBONYD   | 305       | 305                                       | N-LINKED (GLCNAC; . .) (POTENTIAL).       |  |
| FT               | CARBONYD   | 417       | 423                                       | N-LINKED (GLCNAC; . .) (POTENTIAL).       |  |

Query Match 89.0%; Score 1565; DB 1; Length 3011;  
 Best Local Similarity 84.2%; Pred. No. 3; e=124; Mismatches 26; Indels 0; Gaps 0;  
 Matches 287; Conservative 28; N-LINKED (GLCNAC. . .) (POTENTIAL).  
**QY** 1 VRGGDAILITLTCVHAPELIPIDKULLAIFGLPLNLOGITKPYFRAAGLIRACMLV 60  
**Db** 866 VRSGRDAVILMCAVHPTVFDITKLIAVFGPELWILQASLKVYPFVRQDILFALA 925  
**QY** 61 RKAAGGHHYQMAFMKLAUTGTYVYDHJTPQQWAHGIRDALAVAVPVISMEVKLT 120  
**Db** 926 RKGAGGHHYQMVTKLGAITGTYVYDHJTPRQWAHGIRDALAVAVPVISMEVKLT 985  
**QY** 121 WGDTAAACDDIISGIPVSARRGRETLGPADNFEGQGRHLAPITAYSQOTHQGLGII 180  
**Db** 986 WGDTAAACDDIINGLPVSARRGRELLGPADGVSKGRHLAPITAYQQTREGLGII 1045  
**QY** 181 SLGRDKNOVEGEYQVSTATOSLATATCNGVWTVEGAGSKTLSPKGPTOMYNTD 240  
**Db** 1046 SLTGRDKNOVEGEYQVSTATOSLATATCNGVWTVEGAGSKTLSPKGPTOMYNTD 1105  
**QY** 241 QDVQWQAPGARSMTPCGGSLYDVRADIVPVRRGSRGSLSLPSVSYKGSS 300  
**Db** 1106 QDVQWQAPQGSRLTPCTCGSSDLVLYVRADIVPVRRGSGRSLSLPSVSYKGSS 1165  
**QY** 301 GGPLCPSPGHANGVFRAAVCTRGAKAVDFIPVENLETMR 341  
**Db** 1166 GGPLCPAGHAGVIFRAAVCTRGAKAVDFIPVENLETMR 1206

**RESULT 6**

|           |   |      |          |
|-----------|---|------|----------|
| POLG_HCVH | STANDARD;   | PRT; | 3011 AA. |
| ID        | POLG_HCVH   |      |          |
| AC        | P27958;   |      |          |
| DT        | 01-AUG-1992 (Rel. 23, Created)  |      |          |
| DT        | 01-AUG-1992 (Rel. 23, Last sequence update)   |      |          |
| DT        | 10-OCT-2003 (Rel. 42, Last annotation update)   |      |          |
| DE        | Genome polyprotein [Contains: Capsid protein C (Core protein) (P22); Envelope glycoprotein E1 (GP32); Envelope glycoprotein E2 (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21) (EC 3.4.99.-); Protease/heilicase NS3 (P70) (Hepatitis virin) (EC 3.4.21.98); Nonstructural protein NS4a (P4); Nonstructural protein NS4b (P27); Nonstructural protein NS5A (P56); Nonstructural protein NS5B (P66) (P70) (RNA-directed RNA polymerase) (EC 2.7.7.48)]. |      |          |
| OS        | Hepatitis C virus (isolate H) (HCV) (HCV).  |      |          |
| OC        | Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepaciviruses.  |      |          |
| OX        | NCBI_TaxID=1108;  |      |          |
| RN        | [1] SEQUENCE FROM N.A.  |      |          |
| RP        | MEDLINE=9202256; PubMed=1658800;  |      |          |
| RA        | Inchauspe G., Zebedee S., Lee D.H.H., Sugitani M., Nasoff M., Prince A.M.; "Genomic structure of the human prototype strain H of hepatitis C virus: comparison with American and Japanese isolates.", Proc. Natl. Acad. Sci. U.S.A., 88:10292-10296 (1991).   |      |          |
| RT        | X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1192-1657.   |      |          |

**RX** MEDLINE=97331322; PubMed=9187654;  
**RA** Yao N., Hesson T., Cable M., Hong Z., Kwong A.D., Le H.V., Weber P.C., RT "Structure of the hepatitis C virus RNA helicase domain.";  
**RT** Nat. Struct. Biol. 4:463-467(1997).  
**RL** [3]  
**RN** X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 1192-1657.  
**RX** MEDLINE=9815421; PubMed=9493270;  
**RA** Kim J.L., Morgenstern K.A., Griffith J.P., Dwyer M.D., Thomson J.A., Murcko M.A., Lin C., Caron P.R.; "Hepatitis C virus NS3 RNA helicase domain with a bound oligonucleotide: the crystal structure provides insights into the mode of unwinding.";  
**RL** Structure 6:89-100(1998).  
**CC** -!- FUNCTION: PROTEASE NS2 IS RESPONSIBLE FOR THE CLEAVAGE OF NS2-NS3.  
**CC** -!- FUNCTION: PROTEASE NS3 IS RESPONSIBLE FOR THE CLEAVAGE OF NS3-NS4A, NS4A-NS4B, NS4B-NS5A AND NS5A-NS5B.  
**CC** -!- FUNCTION: NS4A FORMS A COMPLEX WITH NS3 AND IS ESSENTIAL FOR THE ACTIVATION OF NS3.  
**CC** -!- FUNCTION: NS5A SEEMS TO HAVE A TRANSCRIPTIONAL ACTIVATORY ROLE.  
**CC** -!- FUNCTION: NS5A IS A RNA-DEPENDENT RNA POLYMERASE THAT PLAYS AN ESSENTIAL ROLE IN THE VIRUS REPLICATION.  
**CC** -!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polyprotein, commonly with Asp or Glu in the P6 position, Cys or Thr in P1 and Ser or Ala in P1'.  
**CC** -!- CATALYTIC ACTIVITY: N nucleoside (RNA) (N).  
**CC** -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: E1 AND E2. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND mRNA.  
**CC** -!- PTM: THE STRUCTURAL PROTEINS C, E1 AND E2 ARE PRODUCED BY PROTEOTIC PROCESSING BY THE HOST SIGNAL PEPTIDASES.  
**CC** -!- SIMILARITY: THE NS2 PROTEASE BELONGS TO PEPTIDASE FAMILY U39.  
**CC** -!- SIMILARITY: THE NS3 PROTEASE BELONGS TO PEPTIDASE FAMILY S29.  
**CC** This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/EMBL.html> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
**CC** EMBL; M67663; BAA4534.1; -  
**CC** PIR; A36814; GNWVCB.  
**DR** PDB; 1A1V; 25-NOV-98.  
**DR** DR; 1A1V; 16-FEB-99.  
**DR** DR; 1A1V; 17-JUN-98.  
**DR** DR; MEROPS; S28.001; --.  
**DR** DR; MEROPS; U39.001; --.  
**DR** DR; TRANSPO; T04155; --.  
**DR** DR; InterPro; IPR009003; CYB\_Ser\_trypsin.  
**DR** DR; InterPro; IPR001410; DEAD.  
**DR** DR; InterPro; IPR02522; HCV capsid.  
**DR** DR; InterPro; IPR02521; HCV\_core.  
**DR** DR; InterPro; IPR02519; HCV\_env.  
**DR** DR; InterPro; IPR02531; HCV\_N51.  
**DR** DR; InterPro; IPR02518; HCV\_NS2.  
**DR** DR; InterPro; IPR00745; HCV\_NS4a.  
**DR** DR; InterPro; IPR001490; HCV\_NS4b.  
**DR** DR; InterPro; IPR002868; HCV\_NS5a.  
**DR** DR; InterPro; IPR00166; HCV\_RdRp.  
**DR** DR; InterPro; IPR00165; Helicase\_C.  
**DR** DR; InterPro; IPR004109; Peptidase\_C29.  
**DR** DR; InterPro; IPR00795; RNA\_Pol1\_DS\_PS.  
**DR** DR; InterPro; IPR00794; RNA\_Pol1\_PSVir.  
**DR** DR; Pfam; PF01543; HCV\_capsid; 1.  
**DR** DR; Pfam; PF01542; HCV\_core; 1.  
**DR** DR; Pfam; PF01539; HCV\_env; 1.  
**DR** DR; Pfam; PF01560; HCV\_NS1; 1.  
**DR** DR; Pfam; PF01538; HCV\_NS2; 1.  
**DR** DR; Pfam; PF01537; HCV\_NS3; 1.  
**DR** DR; Pfam; PF01006; HCV\_NS4a; 1.  
**DR** DR; Pfam; PF01001; HCV\_NS4b; 1.

|    |  |   |          |   |                                |
|----|--|---|----------|---|--------------------------------|
| DR | Pfam; PRO01506; HCV_N5a; 1.  | FT  | TURN     | 1419  | 1420                           |
| DR | Pfam; PRO0271; helicase_C; 1.  | FT  | STRAND   | 1432  | 1436                           |
| DR | Pfam; PRO0998; viral_RNP; 1.   | FT  | TURN     | 1438  | 1439                           |
| DR | ProDom; PD18062; HCV_NS1; 1.   | FT  | STRAND   | 1450  | 1453                           |
| DR | SMART; SM00497; DEXD_C_1.  | FT  | STRAND   | 1456  | 1463                           |
| KW | Polypeptide; Glycoprotein; Transferase; RNA-directed RNA polymerase; | FT  | STRAND   | 1471  | 1478                           |
| KW | Core protein; Coat protein; Serine protease;                         | FT  | STRAND   | 1480  | 1488                           |
| KW | Transmembrane; Nonstructural protein; Hydrolase; Serine protease;    | FT  | HELIx    | 1481  | 1488                           |
| KW | 3D-structure; INIT_MET   | FT  | TURN     | 1489  | 1490                           |
| FT | REMOVED FROM CAPSID PROTEIN C BY THE                                 | FT  | STRAND   | 1497  | 1501                           |
| FT | CELLULAR AMINOPEPTIDASE.   | FT  | STRAND   | 1507  | 1507                           |
| FT | CASPID PROTEIN C.  | FT  | STRAND   | 1511  | 1511                           |
| FT | ENVLOPE GLYCOPROTEIN_E1.   | FT  | HELIx    | 1514  | 1527                           |
| FT | ENVLOPE GLYCOPROTEIN_E2.   | FT  | HELIx    | 1532  | 1544                           |
| FT | PROTEIN_P7.  | FT  | STRAND   | 1550  | 1550                           |
| FT | NONSTRUCTURAL PROTEIN_NS2.   | FT  | HELIx    | 1555  | 1564                           |
| FT | PROTEASE_Helicase_NS3.   | FT  | HELIx    | 1570  | 1578                           |
| FT | NONSTRUCTURAL PROTEIN_NS4A.  | FT  | TURN     | 1579  | 1580                           |
| FT | NONSTRUCTURAL PROTEIN_NS4B.  | FT  | HELIx    | 1584  | 1597                           |
| FT | NONSTRUCTURAL PROTEIN_NS5A.  | FT  | TURN     | 1598  | 1598                           |
| FT | NONSTRUCTURAL PROTEIN_NS5B.  | FT  | HELIx    | 1606  | 1611                           |
| FT | NONSTRUCTURAL PROTEIN_NS5B.  | FT  | TURN     | 1614  | 1618                           |
| FT | CHARGE RELAY SYSTEM (BY SIMILARITY).                                 | FT  | STRAND   | 1622  | 1623                           |
| FT | CHARGE RELAY SYSTEM (BY SIMILARITY).                                 | FT  | STRAND   | 1627  | 1627                           |
| FT | CHARGE RELAY SYSTEM (BY SIMILARITY).                                 | FT  | STRAND   | 1635  | 1636                           |
| FT | ATP (POTENTIAL).   | FT  | HELIx    | 1640  | 1652                           |
| FT | DECH BOX.  | SQ  | SEQUENCE | 3011 AA;  | 327142 MW;                     |
| FT | N-LINKED (GLCNAC. . .)   | Query   | Match    | 87.5%   | Score 1555; DB 1; Length 3011; |
| FT | (POTENTIAL).   | Best Local Similarity   | 83.9%    | Pred. No. 2.3e-13;  |                                |
| FT | CARBOHYD   | Matches   | 286;     | Conservative  |                                |
| FT | 234  | Indels  | 0;       | Mismatches 29;  |                                |
| FT | 305  | Gaps  | 0;       | Indels 0;   |                                |
| FT | CARBOHYD   | Db  | 866      | VRGGRDAAVLLTCVPHALFDITKULLAIFGPNLMLAGITKPYFRAQGLIRACMV 60     |                                |
| FT | 417  | QY  | 1        | VRGGGRDAAVLLTCVPHALFDITKULLAIFGPNLMLAGITKPYFRAQGLIRACMV 60    |                                |
| FT | CARBOHYD   | Db  | 926      | RKLASHQYVQMLAIIKGALTGTCTVNHFLAPRDMAHNGRLAVELVIFSDMEVKIT 120   |                                |
| FT | 423  | QY  | 61       | RKAAGHGYVQMFKAIALGALTGTCTVNHFLAPRDMAHNGRLAVELVIFSDMEVKIT 120  |                                |
| FT | CARBOHYD   | Db  | 986      | WGAATRACGDIINGLPVSARGQELIIGFADGMWSKGWHLAPITAVAOQTGIGCIT 1045  |                                |
| FT | 430  | QY  | 121      | WGADTAAGDIISGPVSARSGRETTLGPDNFBQGWLAPITAYAQQTGRLGICIT 180     |                                |
| FT | 448  | QY  | 241      | QDLVGMQAPPGARSMPCTCGSDPLXLYVIRHADIVPVERRGDRGSLSPRPSYKSS 300   |                                |
| FT | 476  | QY  | 1106     | QDLVGMQAPPGARSMPCTCGSDPLXLYVIRHADIVPVERRGDRGSLSPRPSYKSS 1165  |                                |
| FT | CARBOHYD   | QY  | 181      | SLTGRDKNQTEGEVYVNSTATOSFLATCTGCVWTFHAGSKTLAGRKGPITOMTYND 240  |                                |
| FT | 522  | QY  | 1046     | SLTGRDKNQTEGEVYVNSTATOSFLATCTGCVWTFHAGSKTLAGRKGPITOMTYND 1105 |                                |
| FT | 540  | QY  | 241      | QDLVGMQAPPGARSMPCTCGSDPLXLYVIRHADIVPVERRGDRGSLSPRPSYKSS 300   |                                |
| FT | 556  | QY  | 301      | GGPLIQLPSGHAVGTFRAAVCTRGVAKAVDPIPTSMETMR 341                  |                                |
| FT | 576  | QY  | 1166     | GGPLIQLPSGHAVGTFRAAVCTRGVAKAVDPIPTSMETMR 1206                 |                                |
| FT | N-LINKED (GLCNAC. . .)   | DB  |          |   |                                |
| FT | (POTENTIAL).   | DB  |          |   |                                |
| FT | N-LINKED (GLCNAC. . .)   | DB  |          |   |                                |
| FT | (POTENTIAL).   | DB  |          |   |                                |
| FT | TURN   | RESULT ?  |          |   |                                |
| FT | 1232   | POLG_HCVJ6  |          |   |                                |
| FT | 1236   | ID - POLG_HCVJ6   |          |   |                                |
| FT | 1239   | STANDARD;   |          |   |                                |
| FT | 1247   | PRT; 3033 AA.   |          |   |                                |
| FT | 1251   | AC P26650;  |          |   |                                |
| FT | 1258   | DT 01-AUG-1992 (Rel. 23, Created)   |          |   |                                |
| FT | 1272   | DT 01-AUG-1992 (Rel. 23, last sequence update)  |          |   |                                |
| FT | 1277   | DT 28-FEB-2003 (Rel. 41, last annotation update)  |          |   |                                |
| FT | 1281   | DE Genome polyprotein [Contains: Capsid Protein C (Core protein) (P22); Envelope Glycoprotein_E1 (GP32) (GP5); Envelope Glycoprotein_E2 (GP68) (GP70) (NS1); Protein_P7; Nonstructural protein_NS2 (P21)] |          |   |                                |
| FT | 1283   | DE (EC 3.4.22.-); Protease/helicase (P70) (Hepacivirus)   |          |   |                                |
| FT | 1285   | DE (EC 3.4.21.98); Nonstructural protein_NS4A (P41); Nonstructural protein_NS4B (P27); Nonstructural protein_NS5A (P56); Nonstructural protein_NS5B (P66) (RNA-directed RNA polymerase) (EC 2.7.7.48)].   |          |   |                                |
| FT | 1291   | DE Hepatitis C virus (isolate HC-J6) (HCV).   |          |   |                                |
| FT | 1295   | OS Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  |          |   |                                |
| FT | 1301   |   |          |   |                                |
| FT | 1302   |   |          |   |                                |
| FT | 1312   |   |          |   |                                |
| FT | 1317   |   |          |   |                                |
| FT | 1323   |   |          |   |                                |
| FT | 1335   |   |          |   |                                |
| FT | 1340   |   |          |   |                                |
| FT | 1343   |   |          |   |                                |
| FT | 1347   |   |          |   |                                |
| FT | 1352   |   |          |   |                                |
| FT | 1353   |   |          |   |                                |
| FT | 1360   |   |          |   |                                |
| FT | 1366   |   |          |   |                                |
| FT | 1368   |   |          |   |                                |
| FT | 1373   |   |          |   |                                |
| FT | 1375   |   |          |   |                                |
| FT | 1376   |   |          |   |                                |
| FT | 1378   |   |          |   |                                |
| FT | 1380   |   |          |   |                                |
| FT | 1382   |   |          |   |                                |
| FT | 1385   |   |          |   |                                |
| FT | 1389   |   |          |   |                                |
| FT | 1393   |   |          |   |                                |
| FT | 1397   |   |          |   |                                |
| FT | 1409   |   |          |   |                                |
| FT | 1410   |   |          |   |                                |
| FT | 1417   |   |          |   |                                |

OC Hepacivirus.  
 OC NCBI\_TaxID=11113;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE:9204440; PubMed=1658196;  
 RA Okamoto H., Okada S.-I., Sugiyama Y., Kurai K., Lizuka H.,  
 RA Machida A., Miyakawa Y., Matsumi M.;  
 RT "Nucleotide sequence of the genomic RNA of hepatitis C virus isolated  
 from a human carrier: comparison with reported isolates for conserved  
 RT and divergent regions.";  
 RL J. Gen. Virol. 72:2697-2704(1991).  
 CC --!- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are  
 hydrophobic, suggesting a possible membrane-related function. NS3  
 and NS5 may play a role in the viral RNA replication.  
 CC CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral  
 precursor polyprotein, commonly with Asp or Glu in the P6  
 position Cys or Thr in P1 and Ser or Ala in P1'.  
 CC --!- CATALYTIC ACTIVITY: N nucleoside triphosphate +  
 {RNA} (N).  
 CC --!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a  
 lipoprotein envelope. The envelope consists of two proteins:  
 CC protein M and Glycoprotein E. The nucleocapsid is a complex of  
 protein C and mRNA.  
 CC --!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to license@isb-sib.ch).  
 CC

DR EMBL; D00944; BAA0792.1; -.  
 DR ITR; JQ1303; JQ1303.  
 DR HSSP; P27958; IHER.  
 DR MEROPS; S29.001; -.  
 DR InterPro; IPR009003; Cys\_Ser\_trypsin.  
 DR InterPro; IPR0110; DEBD.  
 DR InterPro; IPR0232; HCV\_capsid.  
 DR InterPro; IPR0251; HCV\_core.  
 DR InterPro; IPR0259; HCV\_env.  
 DR InterPro; IPR02531; HCV\_NS1.  
 DR InterPro; IPR02518; HCV\_NS2.  
 DR InterPro; IPR00745; HCV\_NS3a.  
 DR InterPro; IPR00268; HCV\_NS3b.  
 DR InterPro; IPR02166; HCV\_RdRp.  
 DR InterPro; IPR01150; Helicase\_C.  
 DR InterPro; IPR004109; Peptidase\_C29.  
 DR InterPro; IPR00705; RNA\_poli\_Ds\_Ps.  
 DR InterPro; IPR00704; RNA\_poli\_PsVir.  
 DR Pfam; PRO1543; HCV\_capsid; 1.  
 DR Pfam; PRO142; HCV\_core; 1.  
 DR Pfam; PRO1539; HCV\_env; 1.  
 DR Pfam; PRO1560; HCV\_NS1; 1.  
 DR Pfam; PRO1538; HCV\_NS2; 1.  
 DR Pfam; PRO1507; HCV\_NS3; 1.  
 DR Pfam; PRO1006; HCV\_NS4a; 1.  
 DR Pfam; PRO1001; HCV\_NS4b; 1.  
 DR Pfam; PRO1506; HCV\_NS5a; 1.  
 DR Pfam; PRO0271; helicase\_C; 1.  
 DR Pfam; PRO0998; viral\_RdRp; 1.  
 DR PRODom; PDD8062; HCV\_NS1; 1.  
 DR SMART; SM00487; DXDC; 1.  
 KW Polyprotein; Glycoprotein; Transferase; RNA-directed RNA polymerase;  
 KW Core protein; Coat protein; Envelope protein; Helicase; ATP-binding;  
 KW Transmembrane; Nonstructural protein; Hydrolase; Serine protease.  
 FT INIT\_MET 1 1  
 FT CHAIN 1 115  
 FT CHAIN 1 191

Query Match 71.4%; Score 1269; DB 1; Length 3033;  
 Best Local Similarity 67.2%; Pred. No. 4.2e-99; Mismatches 63; Indels 0; Gaps 0;

Matches 229; Conservative 49; Mismatches 63; Indels 0; Gaps 0;

QY ||||| VRRGRDAILLTCAVHPELIDITKULLATFGPLMVLQAGITKVPYFVRAGLIRACMLV 60  
 DB 870 VRRGRDAILLTCAVHPELIDITKULLATFGPLMVLQAGITKVPYFVRAGLIRACMLV 929  
 QY 61 RKGAGGHVQMAFMKLAALTGTYVPHLTPQDWAHAGLROLAVAVEPVIFPSDEMEKII 120  
 DB 930 RHLAGGRVQWMLLAFGRWIGTYVPHLTPMSDWANGLADLAVAPEPPIFSPMEKKIV 989

QY ||||| WGDATTAAGGDTISGLPVSAARGRETELLGPAINDPFGQGWRLLAPITAYQQTORGGLGICIT 180  
 DB 990 WGAETAAQGDIHGLPVSAARGRETELLGPAINDPFGQGWRLLAPITAYQQTORGGLGICIT 1049

QY 181 SLTGRDKRNQVEGVQTVSTAQSFATCATVAGCWCWVPHGAGSKTLAGPKPQTOMTNVD 240  
 DB 1050 SMTGRDKRNQVEGVQTVSTAQSFATCATVAGCWCWVPHGAGSKTLAGPKPQTOMTNVD 1109

QY 241 QDLVQAPPPGARSMPTCGTSSDYLVTIHADVLPVRREGDSRSLSRPPVSLKGSS 300  
 DB 1110 GDLVQAPPPGARSMPTCGTSSDYLVTIHADVLPVRREGDSRSLSRPPVSLKGSS 1169

QY 301 CGPLPSGHAGVAVCVRGVAKAVDIPVEMTBIR 341  
 DB 1170 GGPVLCPPRGHAVGVFRAAVCSRGVAKSDFIPVETLDIVR 1210

RESULT 8

POLG\_HCVJ8 ID POLG\_HCVJ8 STANDARD; PRT; 3033 AA.

AC P26661; DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)

DB Genome Polyprotein [Contains: Capsid Protein C (Core protein) (P22);  
 DE Envelope glycoprotein E1 (GP2) (GP35); Envelope glycoprotein E2  
 DE (GP68) (GP70) (NS1); Protein P7; Nonstructural protein NS2 (P21)  
 DE (EC 3.4.22.-); Protease/helicase NS3 (P70) (Hepacivirus)]

(EC 3.4.21.98); Nonstructural protein NS4A (P4); Nonstructural protein NS4B (P2'); Nonstructural protein NS5A (P5); Nonstructural protein NS5B (P6) (RNA-directed RNA polymerase) (EC 2.7.7.48)].

DE Hepatitis C virus (isolate HC-J8) (HCV).  
 OS Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepacivirus;  
 OC NCBI\_TaxID=11115;  
 OX [1];

RN SEQUENCE FROM N.A.  
 RP MEDLINE-92230232; RX  
 RA Okamoto H., Kurai K., Okada S.-I., Yamamoto K., Lizuka H., Tanaka T.,  
 RA Fukuda S., Truda F., Mishiro S.;  
 RT "Full-length sequence of a hepatitis C virus genome having poor  
 homology to reported isolates: comparative study of four distinct  
 genotypes.";  
 RL virology 188:331-341(1992).

-!- FUNCTION: The small proteins NS2A, NS2B, NS4A and NS4B are hydrophobic, suggesting a possible membrane-related function. NS3 and NS5 may play a role in the viral RNA replication.

-!- CATALYTIC ACTIVITY: Hydrolysis of four peptide bonds in the viral precursor polyprotein, commonly with Asp or Glu in the P6 position. Cys or Thr in P1 and Ser or Ala in P1'.  
 -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + (RNA)(N).

-!- SUBUNIT: The virion of this virus is a nucleocapsid covered by a lipoprotein envelope. The envelope consists of two proteins: protein M and Glycoprotein E. The nucleocapsid is a complex of protein C and MNA.

-!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY S29.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.iab-sib.ch/announce/> or send an email to license@sb-sib.ch).

---

CC EMBL: D|0988; BAA0761; 1; .

DR PIR: A4050; GNWVB8.

DR HSSP; P21958; IHET; .

DR MEROPS; S29; 001; .

DR InterPro; IPR00903; Cys\_Ser\_trypsin.

DR InterPro; IPR01410; DEAD.

DR InterPro; IPR02522; HCV\_capsid.

DR InterPro; IPR02523; HCV\_core.

DR InterPro; IPR002519; HCV\_env.

DR InterPro; IPR002531; HCV\_NS1.

DR InterPro; IPR002518; HCV\_NS2.

DR InterPro; IPR00745; HCV\_NS4a.

DR InterPro; IPR001490; HCV\_NS4b.

DR InterPro; IPR002868; HCV\_NS5a.

DR InterPro; IPR002165; HCV\_RdRp.

DR InterPro; IPR004109; Peptidase\_C29.

DR InterPro; IPR007095; RNA\_pol\_DS\_PS.

DR InterPro; IPR007094; RNA\_PoI\_PSVir.

DR Pfam; PF01543; HCV\_capsid; 1.

DR Pfam; PF01542; HCV\_core; 1.

DR Pfam; PF01539; HCV\_env; 1.

DR Pfam; PF01560; HCV\_NS1; 1.

DR Pfam; PF01538; HCV\_NS2; 1.

DR Pfam; PF02907; HCV\_NS3; 1.

DR Pfam; PF01006; HCV\_NS4a; 1.

DR Pfam; PF01001; HCV\_NS4b; 1.

DR Pfam; PF01506; HCV\_NS5a; 1.

DR Pfam; PF00939; Viral\_RdRp; 1.

DR Prodom; P018602; HCV\_NS1; 1.

DR SMART; SM00487; DEPDc; 1.

DR SMRT; SM00487; DEPDc; 1.

KW Polyprotein; Glycoprotein; Transferase: RNA-directed RNA polymerase; Core protein; Coat protein; Envelope protein; Helicase; ATP-binding; Transmembrane; Nonstructural protein; Hydrolase; Serine protease.

INIT\_MET 1 REMOVED FROM CAPSID PROTEIN C BY THE

| FT                    | PT       | CHAIN   | 1        | 115               | CELLULAR AMINOPEPTIDASE.                 |
|-----------------------|----------|---|----------|-------------------|--|
| FT                    | PT       | CHAIN   | 116      | 191               | CAPSID PROTEIN C (POTENTIAL).            |
| FT                    | PT       | CHAIN   | 192      | 383               | MAJOR ENVELOPE PROTEIN (POTENTIAL).      |
| FT                    | PT       | CHAIN   | 384      | 733               | NONSTRUCTURAL PROTEIN NS1 (POTENTIAL).   |
| FT                    | PT       | CHAIN   | 734      | 1010              | NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).   |
| FT                    | PT       | CHAIN   | 1011     | 1619              | PROTEASE/HELICASE NS3 (POTENTIAL).       |
| FT                    | PT       | CHAIN   | 1620     | 1866              | NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).  |
| FT                    | PT       | CHAIN   | 1867     | 2017              | NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).  |
| FT                    | PT       | CHAIN   | 2018     | 3033              | RNA-DIRECTED RNA POLYMERASE (POTENTIAL). |
| FT                    | PT       | TRANSMEM  | 347      | 369               | POTENTIAL.                               |
| FT                    | PT       | ACT SITE  | 1087     | 1087              | CHANGE RELAY SYSTEM (BY SIMILARITY).     |
| FT                    | PT       | ACT SITE  | 1111     | 1111              | CHANGE RELAY SYSTEM (BY SIMILARITY).     |
| FT                    | PT       | ACT SITE  | 1169     | 1169              | CHANGE RELAY SYSTEM (BY SIMILARITY).     |
| FT                    | PT       | NP BIND   | 1234     | 1241              | ATP (POTENTIAL).                         |
| FT                    | PT       | SITE  | 1320     | 1323              | DECH BOX.                                |
| FT                    | PT       | CARBONYD  | 196      | 196               | N-LINKED (GLCNAC. . .) (POTENTIAL).      |
| FT                    | PT       | CARBONYD  | 209      | 209               | N-LINKED (GLCNAC. . .) (POTENTIAL).      |
| FT                    | PT       | CARBONYD  | 233      | 233               | N-LINKED (GLCNAC. . .) (POTENTIAL).      |
| FT                    | PT       | CARBONYD  | 299      | 299               | N-LINKED (GLCNAC. . .) (POTENTIAL).      |
| FT                    | PT       | CARBONYD  | 305      | 305               | N-LINKED (GLCNAC. . .) (POTENTIAL).      |
| FT                    | PT       | CARBONYD  | 417      | 417               | N-LINKED (GLCNAC. . .) (POTENTIAL).      |
| FT                    | PT       | CARBONYD  | 423      | 423               | N-LINKED (GLCNAC. . .) (POTENTIAL).      |
| FT                    | PT       | CARBONYD  | 430      | 430               | N-LINKED (GLCNAC. . .) (POTENTIAL).      |
| FT                    | PT       | CARBONYD  | 448      | 448               | N-LINKED (GLCNAC. . .) (POTENTIAL).      |
| FT                    | PT       | CARBONYD  | 477      | 477               | N-LINKED (GLCNAC. . .) (POTENTIAL).      |
| FT                    | PT       | CARBONYD  | 534      | 534               | N-LINKED (GLCNAC. . .) (POTENTIAL).      |
| FT                    | PT       | CARBONYD  | 542      | 542               | N-LINKED (GLCNAC. . .) (POTENTIAL).      |
| FT                    | PT       | CARBONYD  | 558      | 558               | N-LINKED (GLCNAC. . .) (POTENTIAL).      |
| FT                    | PT       | CARBONYD  | 578      | 578               | N-LINKED (GLCNAC. . .) (POTENTIAL).      |
| FT                    | PT       | CARBONYD  | 627      | 627               | N-LINKED (GLCNAC. . .) (POTENTIAL).      |
| FT                    | PT       | CARBONYD  | 649      | 649               | N-LINKED (GLCNAC. . .) (POTENTIAL).      |
| FT                    | PT       | CARBONYD  | 1091     | 1091              | N-LINKED (GLCNAC. . .) (POTENTIAL).      |
| FT                    | PT       | CARBONYD  | 2038     | 2038              | N-LINKED (GLCNAC. . .) (POTENTIAL).      |
| FT                    | PT       | CARBONYD  | 2359     | 2359              | N-LINKED (GLCNAC. . .) (POTENTIAL).      |
| FT                    | PT       | CARBONYD  | 2811     | 2811              | N-LINKED (GLCNAC. . .) (POTENTIAL).      |
| SO                    | SEQUENCE |   | 3033 AA; | 33017 MW;         | LA173E7E3381FDIA CRC64;                  |
| Query                 | Match    |   | 71.0%    | Score 1265;       | DB 1;                                    |
| Best Local Similarity |          |   | 56.0%    | Pred. No. 1. 6.6- | Length 3033;                             |
| Matches               |          |   | 54;      | Mismatches 62;    | Indels 0;                                |
| Conservative          |          |   |          |                   | Gaps 0;                                  |
| OY                    | 1        | VRGGRDAILLTCAVHPLEPLDFITKLIAFGPLVLAQITKPYFVRAQGLTRACKLV           | 60       |                   |  |
| OY                    | 870      | VRGGRRGIIWAVILHPRVLFEVTKMILATIGPVLKIKLAPSLTAAQGLDLA               | 929      |                   |  |
| OY                    | 61       | RKAAGGHYVQAFMKLALGTGTYVDLTLBQLDWAHAGRDLLAVAPEPVFSNEVKIT           | 120      |                   |  |
| OY                    | :        | :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   : |          |                   |  |
| Db                    | 930      | KHLAGARVIVQMLLITGRWTGIVDHLISLSTWAQGLDLA                           | 989      |                   |  |
| OY                    | 121      | WGADTAAGDITISGLVSAARGRETLIGRDNFEGCGWLLAPITAYSSQORTGIGCIT          | 180      |                   |  |
| OY                    | :        | :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   : |          |                   |  |
| Db                    | 990      | WGAETVAGDPLIGLPGVSARLGREVLGPGADYTSKKGWLKAPITAYTQOTRGIGAIV         | 1049     |                   |  |
| OY                    | 181      | SLTRGRKQNGVEGVWSTATOSFLATCYNGVCMWVFHAGSKTLAGPKGPITOMYTND          | 240      |                   |  |
| Db                    | 1050     | SLTRGRKQNEQAGQVOVQVLSVNTQFLGTSISGLVWVYHGAGNKTLAGPKGPVQMTSAE       | 1109     |                   |  |
| OY                    | 241      | QDLVGKQAPPGRASMPCTCSDYLVTRHADVPVRRGDSRGSLSLSPRPSVYLGSS            | 300      |                   |  |
| Db                    | 1110     | GDLVGKQSPKTSPLCTGAVDLYVLTENADVIPVRRKDQRGALISPRPLSTLGSS            | 1169     |                   |  |
| OY                    | 301      | CGPLLCESGHAVGIFPAVTRGVAKVADTIPVSEMEITR                            | 341      |                   |  |
| Db                    | 1170     | GDPVLCSRGAHVGLFRAVCAVGVAKSIDFIPVSELDVATR                          | 1210     |                   |  |

RESULT 9

VET2 REVBU ID VST2 REVBU STANDARD; PRT; 660 AA.  
 AC P29375; DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)

DT 01-FEB-1994 (Rel. 28, Last annotation update)  
 DE Structural protein 2 precursor (ORF2).  
 OS Hepatitis E virus (strain Burma) (HEV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage;  
 OC Hepatitis E-like viruses.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=202467; PubMed=1926770;  
 RA Tam A.W., Smith M.M., Guerra M.E., Huang C.-C., Bradley D.W., Fry K.E., Reyes G.R.,  
 RT "Hepatitis E virus (HEV): molecular cloning and sequencing of the full-length viral genome.",  
 RL Virology 185:120-131(1991).  
 CC - FUNCTION: CONTAINS A HIGH BASIC AMINO ACID CONTENT SUGGESTING THAT IT MAY BE INVOLVED IN THE ENCAPSIDATION OF THE GENOMIC RNA BY EFFECTIVELY NEUTRALIZING THE NEGATIVELY CHARGED RNA.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement. (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; MT3218; AAA45736.1; -.  
 DR PIR; C40778; VHHWH2.  
 DR InterPro; IPR004221; SP2.  
 DR InterPro; IPR008975; Viral\_cap\_coat.  
 DR Pfam; PF03014; SP2; 1.  
 KW Signal.

FT SIGNAL 1  
 FT CHAIN 20 660 AA; 70978 MW; 5832A013CCCCA61C CRC64;  
 SQ SEQUENCE 660 AA; 70978 MW; 5832A013CCCCA61C CRC64;

Query Match 5.8%; Score 102.5; DB 1; Length 660;  
 Best Local Similarity 19.8%; Pred. No. 0.55; Mismatches 130; Indels 151; Gaps 19; Matches 82; Conservative.

QY 24 TKLLI--AIFGPLMVHQAG---ITKVPYFVR--AQGLIRACMLVKRAAGGHYQMA 72  
 Db 151 TNLYVLAAPSLPLQDGTNTHIMATEASNYAQYRVARATIRYRPLVNAVGGAIS 210  
 QY 73 FMKLAALTGTYVYHILPTLQDWAHAGLRLDAVAVEPVPIFSDMEVKILTGWADTAACDDII 132  
 Db 211 FWPQTTPITSV-----DMNSITSTDVRLVQPGIASELVI----- 246  
 QY 133 SGLPVSARRGRREILIGPAD--NFEGQWRILLAPI-TAVSQQRGLL----GCITSLTG 184  
 Db 247 -----PSELRHLHRNQGRSVEETSGVAEEBATSGLVMLCIRHSVSYTN 290  
 QY 185 -----RDKNQVEGVQVSTATOSFL 205  
 Db 291 TPTYGALGILDFALELEFRNLITPGNTNTVRSYSSSTARHRLRGADGTTAAFRM 350  
 QY 206 A----TCVNGV-----CWTVPH-----GAG----SKTLAPKG-PIT 233  
 Db 351 KDLYFTSTINGVGEIGRGLAITLFLADTLIGLPLTELISSAGOLFYSRPVVSANGEPTV 410  
 QY 234 QMTINNDQDLVGMWAPPGRARSMPCTCGSSDLYIV--TRHADVIPTVRRGDSRG-SILLS 289  
 Db 411 KLYTSVEMA---QDKGKAIAPHIDLGESRSRVIQDYNQHQDRPTPPSPSRPSVLR 466  
 QY 290 PRPVSYLK-----GSSGCPPLCPSGHAGVTFRAAVCTRGVAKAVDPFPV 333  
 Db 467 ANDVWLISITAAYDQSTYGSCTGPVV--SDSVLTVNATGAQAVARSIDWTKV 519

RESULT 10  
 VST2\_HVUPA ID\_VST2\_HEVPA STANDARD; PRT; 660 AA.  
 AC P33426;

DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DE Structural protein 2 precursor (ORF2).  
 OS Hepatitis E virus (strain Pakistan) (HEV).  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage;  
 OC Hepatitis E-like viruses.  
 OX NCBI\_TaxID=31767;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92115700; PubMed=1731327;  
 RA Tsarev S.A., Emerson S.U., Reyes G.R., Tsareva T.S., Legters L.J.,  
 RT Malik I.A., Igbl M., Purcell R.H.;  
 RL "Characterization of a prototype strain of hepatitis E virus.", Proc. Natl. Acad. Sci. U.S.A. 89:559-563(1992).  
 CC - FUNCTION: CONTAINS A HIGH BASIC AMINO ACID CONTENT SUGGESTING THAT IT MAY BE INVOLVED IN THE ENCAPSIDATION OF THE GENOMIC RNA BY EFFECTIVELY NEUTRALIZING THE NEGATIVELY CHARGED RNA.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement. (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; MB0081; AAA4527.1; -.  
 DR InterPro; IPR004261; SP2.  
 DR InterPro; IPR008975; Viral\_cap\_coat.  
 DR Pfam; PF03014; SP2; 1.  
 KW Signal.

FT SIGNAL 1  
 FT CHAIN 23 660 AA; 70980 MW; 8085BC53CFB4FB3 CRC64;  
 SQ SEQUENCE 660 AA; 70980 MW; 8085BC53CFB4FB3 CRC64;

Query Match 5.8%; Score 102.5; DB 1; Length 660;  
 Best Local Similarity 19.8%; Pred. No. 0.55; Mismatches 128; Indels 151; Gaps 20; Matches 82; Conservative.

QY 24 TKLLI--AIFGPLMVHQAG---ITKVPYFVR--AQGLIRACMLVKRAAGGHYQMA 72  
 Db 151 TNLYVLAAPSLPLQDGTNTHIMATEASNYAQYRVARATIRYRPLVNAVGGAIS 210  
 QY 73 FMKLAALTGTYVYHILPTLQDWAHAGLRLDAVAVEPVPIFSDMEVKILTGWADTAACDDII 132  
 Db 211 FWPQTTPITSV-----DMNSITSTDVRLVQPGIASELVI----- 246  
 QY 133 SGLPVSARRGRREILIGPAD--NFEGQWRILLAPI-TAVSQQRGLL----GCITSLTG 184  
 Db 247 -----PSELRHLHRNQGRSVEETSGVAEEBATSGLVMLCIRHSVSYTN 290  
 QY 170 QT-RGLGCI-----ITSLIGRDKNO-----VEGEQVNSTATOSFL 205  
 Db 291 TPTYGALGILDFALELEFRNLITPGNTNTVRSYSSSTARHRLRGADGTTAAFRM 350  
 QY 206 A----TCVNGV-----CWTVPH-----GAG----SKTLAPKG-PIT 233  
 Db 351 KDLYFTSTINGVGEIGRGLAITLFLADTLIGLPLTELISSAGOLFYSRPVVSANGEPTV 410  
 QY 234 QMTINNDQDLVGMWAPPGRARSMPCTCGSSDLYIV--TRHADVIPTVRRGDSRG-SILLS 289  
 Db 411 KLYTSVEMA---QDKGKAIAPHIDLGESRSRVIQDYNQHQDRPTPPSPSRPSVLR 466  
 QY 290 PRPVSYLK-----GSSGCPPLCPSGHAGVTFRAAVCTRGVAKAVDPFPV 333  
 Db 467 ANDVWLISITAAYDQSTYGSCTGPVV--SDSVLTVNATGAQAVARSIDWTKV 519

DT DT 01-OCT-1994 (Rel. 30, Created)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Signal recognition particle, chloroplast precursor  
 (SRP54) (54 chloroplast protein) (54CP) (FFC).  
 DE DE OR AT5G09490 OR F8F6 150.  
 GN OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; rosids;  
 OC euroids II; Brassicales; Brassicaceae; Arabidopsids.  
 OC NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv, Columbia;  
 RX MEDLINE:94012817; PubMed=8408079;  
 RA Franklin A.E.; Hoffman N.E.;  
 RT "Characterization of a chloroplast homologue of the 54-kDa subunit of  
 the signal recognition particle";  
 J. Biol. Chem. 268:22175-22180(1993).  
 RL [2]  
 RP SEQUENCE FROM N.A.  
 RA Amin P.; Sy D.; Pilgrim M.; Parry D.H.; Hoffman N.E.;  
 RT "Isolation of two Arabidopsis mutants in the nuclear gene ffc,  
 encoding the 54 kDa subunit of chloroplast signal recognition  
 particle"; Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.  
 RL [3]  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv, Columbia;  
 RX MEDLINE:2101721; PubMed=11130714;  
 RA Tabata S., Kaneko T., Nakamura Y., Kotani H., Kato T., Asamizu E.,  
 Miyajima N., Sasamoto S., Kimura T., Hosouchi T., Kawashima K.,  
 Kohara M., Matsunoto M., Matsuno A., Muraki A., Nakayama S.,  
 Nakazaki N., Naruo K., Okumura S., Shinozaki K., Wada T.,  
 Watanabe A., Yamada M., Yasuda M., Sato S., de la Bastide M.,  
 Huang E., Spiegel L., Gnoj L., O'Shaughnessy A., Preston R.,  
 Habermann K., Murray J., Johnson D., Rohlfing T., Nelson J.,  
 Stoneking T., Pepin K., Spieth J., Sekhon M., Armstrong J., Becker M.,  
 Beiter E., Cordon H., Cordes M., Courtney L., Courtney W., Dante M.,  
 Du H., Edwards J., Fryman J., Haakensen B., Lamar E., Latrelle P.,  
 Leonard S., Meyer R., Mulvaney E., Ozersky P., Riley A., Strokmatt C.,  
 Wagner-McPherson C., Wollam A., Yoakum M., Bell M., Dedina N.,  
 Parmell L., Shah R., Rodriguez M., Hoon See L., Vil D., Baker J.,  
 Kirchoff K., Toth K., King L., Bahret A., Miller B., Marra M.A.,  
 Marrionen R., McCombie W.R., Wilson R.K., Murphy G., Bancroft I.,  
 Volckaert G., Wambott R., Duesterhoeft A., Stikema W., Pohl T.,  
 Entian K.-D., Terry N., Hartley N., Bent E., Johnson S., Zimmermann W.,  
 Langham S. A., McCullagh B., Robben J., Grymonprez B., Zimmerman W.,  
 Raemdonck K., Wedderburn R., Bakke K., Wedderburn R.,  
 Raemdonck K., Wedderburn R., Klein Lankhorst R.,  
 van Staveren M., Dirksen W., Bokma K., Wedderburn R.,  
 Weitzenecker T., Bothe G., Rose M., Hauf J., Bernreiser S., Hempel S.,  
 Feldpausch M., Lambeth S., Villarreal R., Giesen J., Ardiles W.,  
 Bents O., Lemcke K., Kolesov G., Mayer K., Rudd S., Schoof H.,  
 Schueler C., Zaccaria P., Mewes H.-W., Bevan M., Fransz P.,  
 RT "Sequence and analysis of chromosome 5 of the plant Arabidopsis  
 thaliana";  
 RL Nature 418:823-826(2000).  
 CC !- FUNCTION: May target chloroplast proteins to either the thylakoid  
 or envelope membranes.  
 CC !- SUBCELLULAR LOCATION: Chloroplast stroma.  
 CC !- TISSUE SPECIFICITY: Most abundant in green shoot tissue and  
 lower levels seen in the roots and etiolated buds.  
 CC !- SIMILARITY: Belongs to the Grp-binding SRP family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions a long as its content is in no way modified and this statement is not removed. Usage by and/or for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; Z21970; CAA79981.1; -.

DR EMBL; AF092168; AAC64139.1; -.

DR EMBL; AL162873; CAB85514.1; -.

DR PIR; S36637; S36637.

DR HSSP; 007347; 1FFH.

DR InterPro; IPR03593; AAA ATPase.

DR InterPro; IPR00087; SRP54.

DR InterPro; IPR04125; SRP54\_SPB.

DR InterPro; IPR04780; SRP\_sub.

DR Pfam; PF02891; SRP54\_N\_1.

DR Pfam; PF02978; SRP\_SPB; 1.

DR ProDom; PD000819; SRP54; 1.

DR SMART; SM00382; AAA; 1.

DR TIGRFams; TIGR00959; FEF; 1.

DR PROSITE; PS00048; SRP54; 1.

DR Signal recognition particle; GTP-binding; RNA-binding; Chloroplast;

KW Transit Peptide.

RX CHLOROPLAST.

FT SIGNAL RECOGNITION PARTICLE 54 kDa

FT PROTEIN.

FT G-DOMAIN.

FT DOMAIN 76 370

FT DOMAIN 371 564

FT DOMAIN 183 190

FT DOMAIN 265 269

FT DOMAIN 323 326

FT E -> V (IN REF. 2)

SO SEQUENCE 564 AA; 61232 MW; 423F7285FB9063E4 CRC64;

Query Match 5.7%; Score 101; DB 1; Length 564;

Best Local Similarity 26.1%; Pred. No. 0-6; Mismatches 74; Indels 42; Gaps 11; Matches 54; Conservative 37; MisMatches 74; Indels 42; Gaps 11;

QY 15 VHPETI----FDITKULLAFGGLPLVNLTVQGKTVQAKLCYLKKQG--KSCMII- 210

Db 154 VHDELVKLMGEGVSELQFAKSGPTVLLVAGIQLQGVGKTVQAKLCYLKKQG--KSCMII- 210

QY 62 KAGGHVYQMAFMKLAAL---TGYWVPHLPIQ---DWHAHGRLDLYAVEPVIFSDHEV 116

Db 211 --AGDVYRPAPIDQDOLVILGQVGVPVTAQTDVKPADIAKQGKEAKR-----NNNDV 261

QY 117 KLTWGAQDTAACGDTISGLDVSARGREIL---LGAPDNPEGWGRILLAPTAQST 171

Db 262 VIM---DTAGRLQDGKGMDELOVKFLNPTEVLLVVDAMTGQ--RGAALVTFENVI 315

QY 172 RGLIGGICLTSLGRDKNQVGEQVTS 198

Db 316 -GIGGAIKLKDGSRGAAALSVKVS 341

RESULT 12

ID DPO2\_MOUSE STANDARD; PRT; 600 AA.

AC P33611;

DT 01-FEB-1994 (Rel. 28, Created)

DT 01-FEB-1994 (Rel. 28, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE DNA polymerase alpha 70 kDa subunit (DNA polymerase subunit B).

DE POLA2.

OS Mus musculus (Mouse).

OC Buka-rota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OC NCBI\_TaxID=10990;

RN [1]

RP SEQUENCE FROM N.A., AND SEQUENCE OF 84-102; 269-285 AND 394-403.

RX MEDLINE=93116788; PubMed=8463324;

RA Miyazawa H., Izumi M., Tada S., Takada R., Masutani M., Uji M.,

RA Hanacka F.; "Molecular cloning of the cDNAs for the four subunits of mouse DNA polymerase alpha-prime complex and their gene expression during cell proliferation and the cell cycle.";

RL J. Biol. Chem. 268:8111-8122(1993);

CC DR - FUNCTION: May play an essential role at the early stage of chromosomal DNA replication by coupling the polymerase

CC alpha/primate complex to the cellular replication machinery (By similarity).

CC -!- SUBUNIT: DNA polymerase alpha-primate is a four subunit enzyme (subunits A, B, C and D), which is assembled throughout the cell cycle. The largest subunit (subunit A) has DNA polymerase activity; the two smallest subunits (subunits C and D) have DNA primase activity. Subunit B binds to subunit A.

CC -!- SUBCELLULAR LOCATION: Nuclear

CC -!- PTM: PHOSPHORYLATION IN A CELL CYCLE-DEPENDENT MANNER, IN G2/M PHASE (BY SIMILARITY).

CC -!- SIMILARITY: Belongs to the DNA polymerase alpha subunit B family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation of the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC

DR EMBL; DI3546; BAA02746.1; -.

DR PIR; B46642; B46642.

DR MGD; MGI:99690; PG1aa2.

DR InterPro; IPR007200; DNA\_POL\_alpha\_B.

DR Pfam; PF04058; DNA\_Pol\_alpha\_B; 1.

DR SEQUENCE 600 AA; 66267 MW; 79P94B86EF33FEB CRC64;

FT DOMAIN 101 107 PRO/SER/THR-RICH (HYDROPHILIC).

FT DOMAIN 115 157 METAL-BINDING.

FT DOMAIN 600 AA; 66267 MW; 79P94B86EF33FEB CRC64;

Query Match Best Local Similarity 5.7%; Score 101; DB 1; Length 600; Matches 55; Conservative 34; Mismatches 71; Indels 62; Gaps 12;

QY 53 LIRACMLVLRKAAGGHYQVM-AFMKLAALT-----GTIVVDH-----TPLOQWA 95

Db 27 LAELUCLVLYHQTEDGMVSELIAFCTSAGKCOLTDVILNSPEYEVINKLUSKAWHSASKDGS 86

QY 96 HAGIRDLAYAEPVPSDEKVKITWGAQTAACDI--ISGLP-----ISARRGREL 145

Db 87 HAGTRDI-VSIQELIEABEEETIULSYTTPSKGPQLKRSVSTPTEPLTKRSVARSPRQ- 144

QY 146 LGGRADNFEGQWRLAPRTAYSSQTRGLIGCITSLSGRDDKNQVEGVQVNSTATPSL 205

Db 145 LLSPS-----FSPSATPSK-----YTSRINR-----GEVVTFGSAQ--- 178

Qy 205 ATCVNGVCTVIFHGAGSKL-----AGPKGPKITQMYVINDQDLV 245

Db 179 -----GLSWSRGSSGSVSLKVVDPEPLTGSKAMFOOLMG 215

RESULT 13

CYAA—LEIDO STANDARD; PRT; 1380 AA.

ID Q27675; 15-JUL-1998 (Rel. 36, Created)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Receptor-type adenylyl cyclase A (EC 4.6.1.1) (ATP pyrophosphate-lyase) (Adenylyl cyclase).

GN RAC-A.

OS Leishmania donovani.

OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania; OX NCBI\_TaxID=5661;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-1S Sudanese

RX MEDLINE-95340554; Pubmed=7615561;

RA Sanchez M.A., Zeoli D., Klamo E.M., Kavanaugh M.P., Landfear S.M., RT donovanii," J. Biol. Chem. 270:17751-17758(1995).

RL J. Biol. Chem. Could act as a receptor for a unknown ligand.

CC

CC -!- CATALYTIC ACTIVITY: ATP = 3',5'-cyclic AMP + diphosphate.

CC -!- COFACTOR: Birds 1 magnesium ion per subunit (By similarity).

CC -!- SUBCELLULAR LOCATION: Membrane-bound.

CC -!- DEVELOPMENTAL STAGE: Expressed in the insect stage (promastigote) but not in the mammalian host stage of the parasite life cycle.

CC -!- SIMILARITY: Belongs to the adenylyl cyclase class-3 family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation of the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC

DR EMBL; U17042; AA047498.1; -.

DR PIR; T18309; T18309.

DR InterPro; IPR001054; G\_cyclase.

DR Pfam; PF00211; guanylylate\_cyc; 1.

DR SMART; SM00044; CYCG1\_1.

DR PROSITE; PS50128; GUANYLYATE\_CYCLASES\_2; 1.

KW Lyase; CAMP biosynthesis; Transmembrane; Receptor; Glycoprotein; Metal-binding; Magnesium.

FT DOMAIN 1 34 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 35 55 EXTRACELLULAR (POTENTIAL).

FT DOMAIN 913 912 POTENTIAL.

FT DOMAIN 938 938 CYTOPLASMIC (POTENTIAL).

FT METAL 981 981 MAGNESIUM (BY SIMILARITY).

FT CARBOHD 422 422 MAGNETIUM (BY SIMILARITY).

FT CARBOHD 478 478 N-LINKED (GLCNAC. . ) (POTENTIAL).

FT CARBOHD 497 497 N-LINKED (GLCNAC. . ) (POTENTIAL).

FT CARBOHD 567 567 N-LINKED (GLCNAC. . ) (POTENTIAL).

FT SEQUENCE 1380 AA; 151692 MW; 6B2D5F7D3C1107A0 CRC64;

Query Match Best Local Similarity 5.3%; Score 93.5; DB 1; Length 1380; Matches 93; Conservative 51; Mismatches 139; Indels 159; Gaps 24;

Qy 3 GERDAATILLTCAVHP-----ELIFDTKL-----LLAIFGPM-----V 36

Db 101 GGRPIKIL-----HPPDQDNLYDIAEVILHSLARQEKLAVLGLYDGRILTAALSNAADV 155

Qy 37 LQAGITKVP-----YFVRAGQIIRACMLVKAAGHYY-----OMAFWK 75

Db 151 VQSGMLIAPFTGSSGVRTSDSVFVTRAPMVKLVAM-----HIVNLRARRVAFMR 210

Qy 76 LAALTG-----TVVYDHPTQDWAHAGLIRDVLAVA-VEPVIFSDMEYK----- 117

Db 211 -----LTKHAFGBELTYQDUTSL-----LRDPAVLYTPVSSVVEDBEADAMAD 260

Qy 118 -----ITWGADTAAGCDIISGLPWSARRGRERBILIG-----PADNFEGQ 156

Db 261 TNPOQVITWAPVQOYIYFLPKVLPDKRPTTSSAYVIVCSMIQRVVFDVYKLLSAGSIKQ 320

Qy 157 GHLRLLAPITASQQTGRLGCGTTSLGDRKQVGE-----VQVSTATOSFLATCVNG 212

Db 321 DGRILASATT-----SPVSGEGIKYMEVKAQSNVIE---NSG 356

Qy 213 CWTVFHGAGSKTLAGPK-----GPTIQWMTNDQDLVGWOPPGARSMTCTCGSSDLYJ 268

Db 357 SDFYVDDDSRTEITLGSKARBARPSKRK-----VDE---PRVAPHSIAKMLGMSQIIVQ 412

Qy 269 TRHADYVIPRERRGDRGSSLSPRPV---SYLKGGSGGPPLIC-PFGHAVLIFRAAVCTREV 324

Db 413 TLEQTDWIVNRSTYKAGLFQNRFVIGGYVLDGYGP---CEPLAQFLG---ASCYCNQG 468

Qy 325 AKAV-----DIPVSE 337

Db 469 HSSILTVLQNASWDIVPDSSFK 490

| RESULT 14  |   | RESULT 15    |   |
|--|---|--------------|---|
| ID   | TOLB_CHLTH  | ID           | TPE_HORSE   |
| AC   | STANDARD;   | STANDARD;    | STANDARD;   |
| Q8KEQ0;  | PRT; 434 AA.  | PRT; 706 AA. | PRT; 706 AA.  |
| DT   | 10-OCT-2003 (Rel. 42, last sequence update)   | DT           | 01-AUG-1992 (Rel. 23, last sequence update)   |
| DE   | TolB protein precursor.   | DE           | TolB protein precursor.   |
| GN   | Chlorobium tepidum  | GN           | Chlorobium tepidum  |
| OS   | Bacteria; Chlorobi; Chlorobia; Chlorobiales; Chlorobiaceae;   | OS           | Bacteria; Chlorobi; Chlorobia; Chlorobiales; Chlorobiaceae;   |
| OC   | Chlorobium  | OC           | Chlorobium  |
| OX   | NCBI_TAXID=1097;  | OX           | Equus caballus (Horse)  |
| RN   | [1] SEQUENCE FROM N.A.  | RN           | Equus caballus (Horse); Equidae; Equus  |
| RP   | SEQUENCE FROM ATCC 49652 / DSM 12025;   | RP           | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;   |
| RX   | STRAIN=TS / ATCC 49652; PubMed=12033901; MEDLINE=2213655;   | RX           | Gn  |
| Eisen J.A., Nealon M.L., Nelson W.C., Haft D.H., Dodson R.J., Debay R., Gwynn J.L., Kolonay J.L., Yang F., Hickey E.K., Peterson J.D., Durkin A.S., Holt I., Umayan L.A., Mason T., Brenner M., Shear T.P., Parksey D., Nieman W.C., Feldblyum T.V., Hansen C.L., Craven M.B., Radune D., Vamathevan J., Khouri O., Gruber T.M., Ketchum K.A., Ventrelli C.J., Tettelin H., Bryant D.A., Fraser C.M.; [2] "The complete genome sequence of <i>Chlorobium tepidum</i> TS, a photosynthetic, anaerobic, green-sulfur bacterium"; Proc. Natl. Acad. Sci. U.S.A. 99:9509-9514(2002). | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi; Mammalia; Eutheria; Perissodactyla; Equidae; Equus  |              |   |
| RA   | -!- FUNCTION: Involved in the tonB-independent uptake of proteins (BY similarity).  | RA           | Mammalia; Eutheria; Perissodactyla; Equidae; Equus  |
| RA   | -!- SUBCELLULAR LOCATION: Periplasmic (Potential).  | RA           | NBGI_TAXID=9796;  |
| RA   | -!- SIMILARITY: Belongs to the tolB family.   | RA           | SEQUENCE FROM N.A.  |
| RL   | This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to license@isb-sib.ch). | RL           | SEQUENCE FROM N.A.  |
| CC   | RT  | RC           | TISSUE-Extrabryonic tissue;   |
| CC   | CC  | RA           | MDowell K.J., Adam M.H., Baker C.B.; Submitted (MAR-1995) to the EMBL/GenBank/DDBJ databases.   |
| CC   | CC  | RA           | -!- FUNCTION: Transferrins are iron binding transport proteins which can bind two atoms of ferric iron in association with the binding of an anion, usually bicarbonate. It is responsible for the transpor   |
| CC   | CC  | RA           | of iron from sites of absorption and heme degradation to those of storage and utilization. Serum transferrin may also have a further role in stimulating cell proliferation.  |
| CC   | CC  | RA           | -!- SUBUNIT: Monomer.   |
| CC   | CC  | RA           | -!- SUBCELLULAR LOCATION: Secreted.   |
| CC   | CC  | RA           | -!- TISSUE SPECIFICITY: Expressed by the liver and secreted in plasma.  |
| CC   | CC  | RA           | -!- DOMAIN: Composed of two homologous domains.   |
| CC   | CC  | RA           | -!- SIMILARITY: Belongs to the transferrin family.  |
| CC   | CC  | RA           | This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to license@isb-sib.ch). |
| CC   | CC  | DR           | EMBL: MG9020; AAA33083; 1; -.   |
| CC   | CC  | DR           | EMBL: U21127; AAH63684; 1; -.   |
| CC   | CC  | DR           | PIR: S33761; S33761.  |
| CC   | CC  | DR           | HSSP: P02787; 1ABE.   |
| CC   | CC  | DR           | InterPro: IPR001156; Transferrin.   |
| CC   | CC  | DR           | ProDom: PF00405; Transferrin_2.   |
| CC   | CC  | DR           | PRINTS: PRO0422; TRANSPERRIN.   |
| CC   | CC  | DR           | SMART: SNO004; TR_FER_2.  |
| CC   | CC  | DR           | PROSITE: PS00205; TRANSFERRIN_1; 2.   |
| CC   | CC  | DR           | PROSITE: PS00206; TRANSFERRIN_2; 2.   |
| CC   | CC  | DR           | PROSITE: PS00207; TRANSFERRIN_3; 2.   |
| CC   | CC  | DR           | Transport; Iron transport; Glycoprotein; Metal-binding; Repeat;   |
| CC   | CC  | KW           | Transport; Iron transport; Glycoprotein; Metal-binding; Repeat; Signal.   |
| CC   | CC  | FT           | 1 19 BY SIMILARITY.   |
| CC   | CC  | FT           | 1 706 SEROTRANSFERRIN.  |
| CC   | CC  | FT           | 20 357 1.   |
| CC   | CC  | FT           | 20 706 2.   |
| CC   | CC  | FT           | 358 64 BY SIMILARITY.   |
| CC   | CC  | FT           | 26 55 BY SIMILARITY.  |
| CC   | CC  | FT           | 36 55 BY SIMILARITY.  |
| CC   | CC  | FT           | 134 215 BY SIMILARITY.  |
| CC   | CC  | FT           | 174 190 BY SIMILARITY.  |
| CC   | CC  | FT           | 177 198 BY SIMILARITY.  |
| CC   | CC  | FT           | 187 200 BY SIMILARITY.  |
| CC   | CC  | FT           | 248 BY SIMILARITY.  |

FT DISULFID 360 623 BY SIMILARITY.  
 FT DISULFID 366 398 BY SIMILARITY.  
 FT DISULFID 376 389 BY SIMILARITY.  
 FT DISULFID 423 701 BY SIMILARITY.  
 FT DISULFID 441 664 BY SIMILARITY.  
 FT DISULFID 474 550 BY SIMILARITY.  
 FT DISULFID 498 692 BY SIMILARITY.  
 FT DISULFID 508 522 BY SIMILARITY.  
 FT DISULFID 519 533 BY SIMILARITY.  
 FT DISULFID 590 604 BY SIMILARITY.  
 FT DISULFID 642 647 BY SIMILARITY.  
 FT METAL 79 79 BY SIMILARITY.  
 FT METAL 111 111 BY SIMILARITY.  
 FT METAL 209 209 BY SIMILARITY.  
 FT METAL 270 270 BY SIMILARITY.  
 FT METAL 413 413 BY SIMILARITY.  
 FT METAL 449 449 BY SIMILARITY.  
 FT METAL 544 544 BY SIMILARITY.  
 FT METAL 612 612 BY SIMILARITY.  
 FT BINDING 136 136 BY SIMILARITY.  
 FT BINDING 140 140 BY SIMILARITY.  
 FT BINDING 142 142 BY SIMILARITY.  
 FT BINDING 143 143 CARBONATE 1 (VIA AMIDE NITROGEN) (BY SIMILARITY).  
 FT BINDING 476 476 CARBONATE 2 (BY SIMILARITY).  
 FT BINDING 480 480 CARBONATE 2 (VIA AMIDE NITROGEN) (BY SIMILARITY).  
 FT BINDING 482 482 CARBONATE 2 (VIA AMIDE NITROGEN) (BY SIMILARITY).  
 FT BINDING 483 483 CARBONATE 2 (VIA AMIDE NITROGEN) (BY SIMILARITY).  
 FT CARBOHYD 515 515 N-LINKED (GLCNAC . C) (POTENTIAL).  
 SQ SEQUENCE 706 AA; 78094 MW; 1A0FA566C0409D8A CRC64;

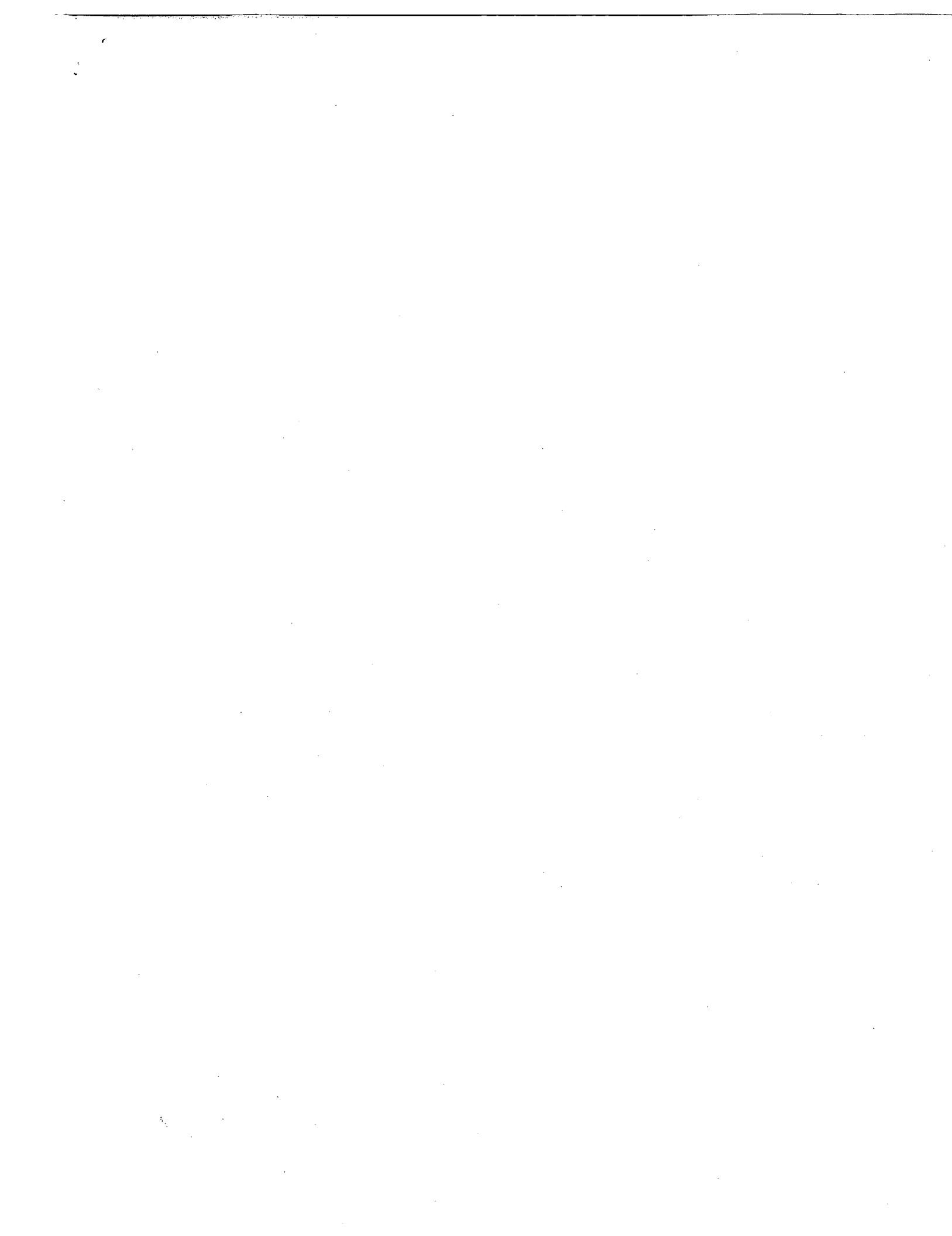
Query Match 5.2%; Score 92.5; DB 1; Length 706;  
 Best Local Similarity 21.5%; Pred. No. 4.2;  
 Matches 64; Conservative 43; Mismatches 109; Indels 81; Gaps 17;

```

  QY 71 MAFMKLAALTGTGTYV--YQHHTPLQDWIAAGLRLAVATEPVVIFSDMVKIITWGA--- 123
  Db 321 IGFPLRIPPANDTWLYLGVYVT-----AIRNLREDIRPEVPRD-ECKKVWCAIGHH 371
  Qy 124 DTAACGD-IIISGLIPVSARRGR-----EILLGPADNEFEGQWRLL-----LAPTA 167
  Db 372 EKVKGDEWSVNSGGNTECESAQSSTEDCTAKIVKGSEADAMSLDGSGFIYTAGKCGILVPLAE 431
  Qy 168 SQQTGGLGCITSITGRDKNQVEGVQVNSTATOSFLATCVNGVCWTVFHGAGSKTLAG 227
  Db 432 NYETRSGSACVDTBEGYH-----AVATVKSSSDPDUT-----W-----NSLKG 470
  Qy 228 PKGPIPTQMVNTDQDVLGQAPPGARSMTPTCGSSDAYLVTRHADIVPVRREGDSRSL 287
  Db 471 KK---SCHTGDR-TAGMNIPMGL-----LYSEKIECFDKFFREGCAGYR 513
  Qy 288 LSPREVPSVLRGSSGCP-LIC-PSGEA----VGFRAAVCTREVAKAVDPIVESME 337
  Db 514 RNSTICNLCTGSASGPGRCEPNHERYVIGTAFRCVLEKDV-----FVXQHIVE 566

```

Search completed: May 6, 2004, 09:31:54  
 Job time : 8.119 secs





DR GO; GO:005199; F:structural molecule activity; IEA.  
 DR GO; GO:0016740; F:transf erase activity; IEA.  
 DR GO; GO:006118; P:electron transport; IEA.  
 DR GO; GO:000508; P:proteolysis and peptidolysis; IEA.  
 DR GO; GO:0006350; P:transcription; IEA.  
 DR GO; GO:0019079; P:viral genome replication; IEA.  
 DR InterPro; IPR009001; Cys\_Ser\_trypsin.  
 DR InterPro; IPR00346; CYC\_heme\_BS.  
 DR InterPro; IPR01410; DEAD.  
 DR InterPro; IPR02522; HCV\_capsid.  
 DR InterPro; IPR02521; HCV\_core.  
 DR InterPro; IPR002519; HCV\_env.  
 DR InterPro; IPR002531; HCV\_NS1.  
 DR InterPro; IPR02538; HCV\_NS2.  
 DR InterPro; IPR00745; HCV\_NS4a.  
 DR InterPro; IPR001430; HCV\_NS4b.  
 DR InterPro; IPR02865; HCV\_NS5a.  
 DR InterPro; IPR002165; HCV\_RdRp.  
 DR InterPro; IPR002160; Helicase\_C.  
 DR InterPro; IPR004109; Peptidase\_C29.  
 DR InterPro; IPR007095; RNA\_Pol\_DS\_Ps.  
 DR InterPro; IPR007094; RNA\_Pol\_Psvir.  
 DR Pfam; PF01001; HCV\_ns4b; 1.  
 DR Pfam; PF01542; HCV\_core; 1.  
 DR Pfam; PF01539; HCV\_env; 1.  
 DR Pfam; PF01560; HCV\_NS1; 1.  
 DR Pfam; PF01538; HCV\_NS2; 1.  
 DR Pfam; PF02907; HCV\_NS3; 1.  
 DR Pfam; PF01005; HCV\_NS4a; 1.  
 DR Pfam; PF01001; HCV\_NS4b; 1.  
 DR Pfam; PF01506; HCV\_NS5a; 1.  
 DR Pfam; PF00271; helicase\_C; 1.  
 DR Pfam; PF00998; viral\_RdRp; 1.  
 DR Prodrom; PD186662; HCV\_NS1; 1.  
 DR SMART; SM00490; HEIIC; 1.  
 DR SMART; SM00487; DEDDC; 1.  
 DR SEQUENCE; 3010 AA; 327042 MW; 3807DC68/9684C95 CRC64;  
 DR SEQUENCE; 3010 AA; 327042 MW; 3807DC68/9684C95 CRC64;

Query Match . . . . . Score 1729; DB 12; Length 3010;  
 Best Local Similarity . . . . . 95.9%; Pred. No. 1; 3e-138;  
 Matches 327; Conservative 8; Mismatches 6; Indels 0; Gaps 0;

QY 1 VRGGDAAILLTCAVHPFLFDIKLLAIFGGLPLMQLAQITKUPYFRAQGLIRACLV 60  
 Db 866 VGRSDAIIILTCAAHPELFIDIKLLAIFGGLPLMQLAQITKUPYFRAQGLIRACLV 925

QY 61 RKAAGHGVVMAFMKAALIGTVYDHALPLODAMAHAGRLDAVAVEVIFSMVEKIT 120  
 Db 926 RKVAGHGVVMAFMKAALIGTVYDHALPLODAMAHAGRLDAVAVEVIFSMVEKIT 985

QY 121 WGADPAACGDIISGLLPVSARGRGTRILGGRADNPGQGRLLAPITATYSQQTGLGCIT 180  
 Db 986 WGADPAACGDIISGLLPVSARGRGTRILGGRADNPGQGRLLAPITATYSQQTGLGCIT 1045

QY 181 SLTGRDKNQVEGEVQVSTATOSLATCTGNGVWTFVFGASKTLSPKGPTOMYNTD 240  
 Db 1046 SLTGRDKNQVEGEVQVSTATOSLATCTGNGVWTFVFGASKTLSPKGPTOMYNTD 1105

QY 241 QDLVQWAQGARGARTMPTCGGSLYLTRHAIPIVERRGDSRGSLSPPRVSYLKSS 300  
 Db 1106 QDLVQWAQGARGARTMPTCGGSLYLTRHAIPIVERRGDSRGSLSPPRVSYLKSS 1165

QY 301 GGPLCPSGHAVGFRANCTGNGVAKAVIDFIPESMETMR 341  
 Db 1166 GGPLCPSGHAVGFRANCTGNGVAKAVIDFIPESMETMR 1206

DR GO; GO:005199; F:structural molecule activity; IEA.  
 DR GO; GO:0016740; F:transf erase activity; IEA.  
 DR GO; GO:006118; P:electron transport; IEA.  
 DR GO; GO:000508; P:proteolysis and peptidolysis; IEA.  
 DR GO; GO:0006350; P:transcription; IEA.  
 DR GO; GO:0019079; P:viral genome replication; IEA.  
 DR GO; GO:0019079; P:viral transformation; IEA.  
 DR GO; GO:005198; F:structural molecule activity; IEA.  
 DR GO; GO:0016740; P:electron transport; IEA.  
 DR GO; GO:0006350; P:proteolysis and peptidolysis; IEA.  
 DR GO; GO:0019079; P:viral genome replication; IEA.  
 DR GO; GO:005198; F:structural molecule activity; IEA.  
 DR InterPro; IPR00345; CYC\_heme\_BS.  
 DR InterPro; IPR01410; DEAD.  
 DR InterPro; IPR02522; HCV\_capsid.  
 DR InterPro; IPR002519; HCV\_env.  
 DR InterPro; IPR002518; HCV\_NS4a.  
 DR InterPro; IPR001745; HCV\_NS4b.  
 DR InterPro; IPR001430; HCV\_NS5a.  
 DR InterPro; IPR002868; HCV\_NS5b.  
 DR InterPro; IPR002166; HCV\_RdRp.  
 DR InterPro; IPR001650; Helicase\_C.  
 DR InterPro; IPR004109; Peptidase\_C29.  
 DR InterPro; IPR007095; RNA\_Pol\_DS\_Ps.  
 DR InterPro; IPR007094; RNA\_Pol\_Psvir.  
 DR Pfam; PF01543; HCV\_capsid; 1.  
 DR Pfam; PF01539; HCV\_core; 1.  
 DR Pfam; PF01560; HCV\_NS1; 1.  
 DR Pfam; PF01538; HCV\_NS2; 1.  
 DR Pfam; PF02907; HCV\_NS3; 1.  
 DR Pfam; PF01005; HCV\_NS4a; 1.  
 DR Pfam; PF01001; HCV\_NS4b; 1.  
 DR Pfam; PF01506; HCV\_NS5a; 1.  
 DR Pfam; PF0271; helicase\_C; 1.  
 DR Pfam; PF00998; viral\_RdRp; 1.  
 DR Prodrom; PD18662; HCV\_NS1; 1.

RESULT 2

DR SMART; SM00487; DEXDC; 1.  
 DR PROSINE; PS00190; CYTOCHROME C; 1.  
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
 Polyprotein; RNA-directed DNA Polymerase; Transferase; Transmembrane.  
 SEQUENCE 3010 AA; 327102 MW; 7162C9DB93ER6B0C7 CRC64;

Query Match 97.1%; Score 1726; DB 12; Length 3010;  
 Best Local Similarity 95.6%; Pred. No. 2.4e-138; Mismatches 5; Indels 0; Gaps 0;  
 Matches 326; Conservative 95.6%; Pred. No. 2.4e-138; Mismatches 5; Indels 0; Gaps 0;

QY 1 VRGGRDAIILTCAVHPLBLIFDITKLLAIFGPMVLIQAGITKVYPYVRAOGLIRACMLV 60  
 866 VRGGRDAIILTCAVHPLBLIFDITKLLAIFGPMVLIQAGITKVYPYVRAOGLIRACMLV 925  
 61 RKGAGGHVQMAEMKLALTTGTVYDILTPLODWAHAGLRLDAVAVPVISSDMEVKIT 120  
 QY 926 RKVAGGHVQMAFMKLGALTGVYDILTPLODWAHAGLRLDAVAVPVISSDMEVKIT 985

121 WGADTAAGGDISGLPYSARRREILLGPADNFEQGWRLLAPIATYSQQTGLGCVT 180  
 986 WGADTAAGGDISGLPYSARRREILLGPADNFEQGWRLLAPIATYSQQTGLGCVT 1045

181 SLTGRDKRNQVEEYQVSTATOSFLATCUNGCVWTVHGAGSKTLAGPKGPIQTOMYINVD 240  
 046 SLTGRDKRNQVEEYQVSTATOSFLATCUNGCVWTVHGAGSKTLAGPKGPIQTOMYINVD 1105

241 QDLVGWQAPPGARMSMPTCGSSDLYLVRHADVIPIRRRSRSRGSLSPRVSYLGSS 300  
 QY 1106 QDLVGWQAPPGARSLTCTCGSSDLYLVRHADVIPIRRRSRSRGSLSPRVSYLGSS 1165

301 GGPLICPSGHAVGIPRAAVCTRGVAKAVDFPVVESMETMR 341  
 Db 1166 GGPLICPSGHAVGIPRAAVCTRGVAKAVDFPVVESMETMR 1206

Db 09DTD6 PRELIMINARY; PRT; 3010 AA.  
 ID 09DTD6:  
 AC 01-MAR-2001 (TREMBLEL; 16, Last sequence update)  
 DT 01-OCT-2003 (TREMBLEL; 25, Last annotation update)  
 DE Repatilis C virus.  
 OS Repatilis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Hepatitis C virus.  
 OC Hepatitis C virus.  
 RN [1] SEQUENCE FROM N.A.  
 RP SIRAINHCVT21; Iwata K., Matsumoto M., Matsumoto H., Nakao K.,  
 RA Takahashi K., Ohta Y., Kanai K., Maruo H., Baba K., Hijikata M.,  
 RA Mishiro S.;  
 RT "Hepatitis C virus (HCV) genotype 1b sequences from fifteen patients with hepatocellular carcinoma: the 'progression score' revisited.";  
 RL Submitted (SPP-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA. BY SIMILARITY.  
 EMBL; AB049101; BAB18814.1; -.

PIR; A61196 A61196  
 DR P0246; PQ0246;  
 DR PS0329; FS0329.

DR HSP; P26663; 1JXP  
 GO; GO:001901; C integral to membrane; IEA.  
 DR GO; GO:0016021; C viral envelope; IEA.  
 DR GO; GO:0005524; F-ATP binding; IEA.  
 GO; GO:0008026; F-ATP dependent helicase activity; IEA..  
 DR GO; GO:0005489; Felección transporter activity; IEA.  
 GO; GO:0016787; F hydrolase activity; IEA.  
 GO; GO:0003233; F-RNA binding; IEA.

DR SMART; SM00487; DEXDC; 1.  
 DR PROSINE; PS00190; CYTOCHROME C; 1.  
 KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;  
 Polyprotein; RNA-directed DNA Polymerase; Transferase; Transmembrane.  
 SEQUENCE 3010 AA; 327102 MW; 7162C9DB93ER6B0C7 CRC64;

Query Match 96.8%; Score 1721; DB 12; Length 3010;  
 Best Local Similarity 95.9%; Pred. No. 6.5e-138; Mismatches 7; Indels 0; Gaps 0;  
 Matches 327; Conservative 95.9%; Pred. No. 6.5e-138; Mismatches 7; Indels 0; Gaps 0;

QY 1 VRGGRDAIILTCAVHPLBLIFDITKLLAIFGPMVLIQAGITKVYPYVRAOGLIRACMLV 60  
 866 VRGGRDAIILTCAVHPLBLIFDITKLLAIFGPMVLIQAGITKVYPYVRAOGLIRACMLV 925  
 61 RKGAGGHVQMAEMKLALTTGTVYDILTPLODWAHAGLRLDAVAVPVISSDMEVKIT 120  
 QY 926 RKVAGGHVQMAFMKLGALTGVYDILTPLODWAHAGLRLDAVAVPVISSDMEVKIT 985

121 WGADTAAGGDISGLPYSARRREILLGPADNFEQGWRLLAPIATYSQQTGLGCVT 180  
 986 WGADTAAGGDISGLPYSARRREILLGPADNFEQGWRLLAPIATYSQQTGLGCVT 1045

181 SLTGRDKRNQVEEYQVSTATOSFLATCUNGCVWTVHGAGSKTLAGPKGPIQTOMYINVD 240  
 1046 SLTGRDKRNQVEEYQVSTATOSFLATCUNGCVWTVHGAGSKTLAGPKGPIQTOMYINVD 1105

241 QDLVGWQAPPGARMSMPTCGSSDLYLVRHADVIPIRRRSRSRGSLSPRVSYLGSS 300  
 QY 1106 QDLVGWQAPPGARSLTCTCGSSDLYLVRHADVIPIRRRSRSRGSLSPRVSYLGSS 1165

301 GGPLICPSGHAVGIPRAAVCTRGVAKAVDFPVVESMETMR 341  
 Db 1166 GGPLICPSGHAVGIPRAAVCTRGVAKAVDFPVVESMETMR 1206

**RESULT 4**

ID Q9JH7; PRELIMINARY; PRT; 3010 AA.

AC Q9JH7; 01-OCT-2000 (TREMBrel. 15, Created)

DT 01-OCT-2000 (TREMBrel. 15, Last sequence update)

DT 01-OCT-2003 (TREMBrel. 25, Last annotation update)

DE Genome Polyprotein.

OS Hepatitis C virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.

NCBI\_TaxID=11103;

[1] SEQUENCE FROM N.A.

STRAIN=M115;

Nagayama K., Kuroaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.; "Characteristics of hepatitis C viral genome associated with disease progression.", Submitted (NOV-1995) to the EMBL/GenBank/DDBJ databases.

PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA. (BY SIMILARITY).

EMBL: AF007756; AAC65946.1; -.

PIR: A61196; A61196.

PIR: PQ0246; PQ0246.

PIR; PQ004; PQ0804.

DR HSSP; P6663; JXPF.

DR GO; GO:0016021; C:integral to membrane; IEA.

DR GO; GO:0019028; C:viral capsid; IEA.

DR GO; GO:0019031; C:viral envelope; IEA.

DR GO; GO:005524; F:ATP binding; IEA.

DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.

DR GO; GO:003723; F:RNA binding; IEA.

DR GO; GO:003968; F:RNA-directed RNA polymerase activity; IEA.

DR GO; GO:005198; F:structural molecule activity; IEA.

DR GO; GO:0016740; F:transferase activity; IEA.

DR GO; GO:0006508; F:electron transport; IEA.

DR GO; GO:0006350; F:transcription; IEA.

DR GO; GO:0019079; F:virual genome replication; IEA.

GO; GO:0019087; F:virual transformation; IEA.

DR InterPro; IPR00903; Cys\_Ser\_trypsin.

DR InterPro; IPR0045; Cytc\_heme\_BS.

DR InterPro; IPR0141; DBD.

DR InterPro; IPR00322; HCV\_heme.

DR InterPro; IPR02511; HCV\_capsid.

DR InterPro; IPR02219; HCV\_env.

DR InterPro; IPR00231; HCV\_NS1.

DR InterPro; IPR02518; HCV\_NS2.

DR InterPro; IPR00745; HCV\_NS3a.

DR InterPro; IPR00190; HCV\_NS4b.

DR InterPro; IPR02868; HCV\_NS5a.

DR InterPro; IPR02166; HCV\_NS5b.

DR InterPro; IPR04109; Peptidase\_C29.

DR InterPro; IPR00795; RNA\_PolyPS.

DR InterPro; IPR00704; RNA\_poli\_PSVir.

DR Pfam; PF01543; HCV\_capsid; 1.

DR Pfam; PF01542; HCV\_core; 1.

DR Pfam; PF01535; HCV\_env; 1.

DR Pfam; PF01560; HCV\_NS1; 1.

DR Pfam; PF01538; HCV\_NS2; 1.

DR Pfam; PF00207; HCV\_NS3; 1.

DR Pfam; PF01006; HCV\_NS4a; 1.

DR Pfam; PF01001; HCV\_NS4b; 1.

DR Pfam; PF01505; HCV\_NS5a; 1.

DR Pfam; PF00998; Viral\_RCP; 1.

DR PROSITE; SM00487; DEXDC\_1.

DR PROSITE; PS00190; CYTOCHROME\_C; 1.

---

**RESULT 5**

ID P90191; PRELIMINARY; PRT; 3010 AA.

AC P90191; 01-MAY-1997 (TREMBrel. 03, Created)

DT 01-MAY-1997 (TREMBrel. 03, Last sequence update)

DT 01-OCT-2003 (TREMBrel. 25, Last annotation update)

DE Genome polyprotein.

OS Hepatitis C virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Hepacivirus.

NCBI\_TaxID=11103;

[1] SEQUENCE FROM N.A.

STRAIN=HCV-1b;

Enomoto N.; Submitted (MAY-1995) to the EMBL/GenBank/DDBJ databases.

[2] SEQUENCE FROM N.A.

RC PRELIMINARY; PRT; 3010 AA.

RX MEDLINE=9534024; PubMed=7542279;

RA Enomoto N., Sakuma I., Asahina Y., Kuroasaki M., Murakami T., Yamamoto C., Izumi N., Marumo F., Sato C.; "Comparison of full-length sequences of interferon-sensitive and resistant hepatitis C virus 1b.", J. Clin. Invest. 95:224-230 (1995)

RL LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA. (BY SIMILARITY).

CC PROTEIN C AND MRNA. (BY SIMILARITY).

DR EMBL; D50482; BAA0073.1; -.

DR PIR; A61196; A61196.

DR PIR; PQ0246; PQ0254; PQ0254; PQ0254.

DR PIR; PQ00504; PQ00804.

DR PIR; PS0029; PS0329.

DR PDB; 1DXW; 1Z-JAN.

DR GO; GO:0016021; C:integral to membrane; IEA.

DR GO; GO:0019028; C:viral capsid; IEA.

DR GO; GO:0019031; C:viral envelope; IEA.

DR GO; GO:0005524; F:ATP binding; IEA.

|   |  |    |   |
|---|--|----|---|
| DR  | GO; GO:0008026; F:ATP dependent helicase activity; IEA.  | Db | 1106 QDLVGWQAPPGARSILTPCTCGSSDLYLVTRHADVIPVRRGDSRGSLLSPRPVSYLKGS 1165 |
| DR  | GO; GO:0003723; F:RNA binding; IEA.  | Ov | 301 GAPIIICPSGHAVGIFRAAVCTRGAVAKAVDFPVESMETWR 341                     |
| DR  | GO; GO:0008336; F:serine-type peptidase activity; IEA.   | Db | 1166 GGPLICPSGHAVGIFRAAVCTRGAVAKAVDFPVESMETWR 1206                    |
| DR  | GO; GO:0005198; F:structural molecule activity; IEA.   |    |   |
| DR  | GO; GO:0016740; F:transferase activity; IEA.   |    |   |
| DR  | GO; GO:0006508; F:proteolysis and peptidolysis; IEA.   |    |   |
| DR  | GO; GO:0006350; F:viral genome replication; IEA.   |    |   |
| DR  | GO; GO:0019079; F:viral transformation; IEA.   |    |   |
| DR  | InterPro; IPR009033; Cys_Ser_trypsin.  |    |   |
| DR  | InterPro; IPR001410; DEAD.   |    |   |
| DR  | InterPro; IPR002532; HCV capsid.   |    |   |
| DR  | InterPro; IPR002511; HCV_core.   |    |   |
| DR  | InterPro; IPR002519; HCV_env.  |    |   |
| DR  | InterPro; IPR002531; HCV_NS1.  |    |   |
| DR  | InterPro; IPR002518; HCV_NS2.  |    |   |
| DR  | InterPro; IPR000145; HCV_NS3a.   |    |   |
| DR  | InterPro; IPR001490; HCV_NS3b.   |    |   |
| DR  | InterPro; IPR002848; HCV_NS4a.   |    |   |
| DR  | InterPro; IPR002166; HCV_RdRp.   |    |   |
| DR  | InterPro; IPR001650; Helicase_C.   |    |   |
| DR  | InterPro; IPR004109; peptidase_C29.  |    |   |
| DR  | InterPro; IPR007035; RNA_Pol_DS_PS.  |    |   |
| DR  | InterPro; IPR007034; RNA_pol_Psvir.  |    |   |
| DR  | Pfam; PF01543; HCV_capsid; 1.  |    |   |
| DR  | Pfam; PF01542; HCV_core; 1.  |    |   |
| DR  | Pfam; PF01506; HCV_NS4b; 1.  |    |   |
| DR  | Pfam; PF00271; helicase_C; 1.  |    |   |
| DR  | Pfam; PF00998; viral_RdRp; 1.  |    |   |
| DR  | ProDom; PDB6062; HCV_NSL; 1.   |    |   |
| DR  | SMART; S00487; DExD.   |    |   |
| KW  | Coat protein; Envelope protein; Glycoprotein; Nonstructural protein; Polypeptide; RNA-directed RNA polymerase; Transferase; Transmembrane. |    |   |
| KW  | PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND MRNA (BY SIMILARITY).   |    |   |
| FT  | CHAIN 1 191 383 E1.  |    |   |
| FT  | CHAIN 1 384 809 E2.  |    |   |
| FT  | CHAIN 1 810 1026 NS2.  |    |   |
| FT  | CHAIN 1 1027 1657 NS3.   |    |   |
| FT  | CHAIN 1 1658 1711 NS4A.  |    |   |
| FT  | CHAIN 1 1712 1972 NS4B.  |    |   |
| FT  | CHAIN 1 1973 2419 NS5A.  |    |   |
| FT  | CHAIN 1 2420 3010 NS5B.  |    |   |
| SQ  | SEQUENCE 3010 AA; 327438 MN; 5F15AC675A0C8268 CRC64;   |    |   |
| <b>Query Match</b>  |  |    |   |
| Best Local Similarity 95.3%; Score 1719; DB 12; Length 3010;  |  |    |   |
| Matches 325; Conservative 11; Mismatches 5; Indels 0; Gaps 0; |  |    |   |
| Qy  | 1 VRGGEDAILLTCAVTHPELFDITKLLIAFGPMVIMVQAGITKPYFRAQGIRTRACMVLV 60   |    |   |
| Db  | 866 VRRGGDAILLTCAVTHPELFDITKLLIAFGPMVIMVQAGITKPYFRAQGIRTRACMVLV 925  |    |   |
| Qy  | 61 RKAAGGHYVQMAFMKLALTGTVYDHPLQDWAHAGRLAVAVWVPSIDMEVKIT 120  |    |   |
| Db  | 926 RKVAGGHYVQMAFMKLALTGTVYDHPLQDWAHAGRLAVAVWVPSIDMEVKIT 985   |    |   |
| Qy  | 121 WGADTAACGDTISGIVPSARRGRELLGADNPFQQGWLLAPITAYSQQTGIGCIT 180   |    |   |
| Db  | 986 WGADTAACGDTISGIVPSARRGRELLGADNPFQQGWLLAPITAYSQQTGIGCIT 1045  |    |   |
| Qy  | 181 SLTGRDKNQYEVGVQVSTATOSFLATCNGVCWTFVHGAGSKTLAGPKPITQMYTN 240  |    |   |
| Db  | 1046 SLTGRDKNQYEVGVQVSTATOSFLATCNGVCWTFVHGAGSKTLAGPKPITQMYTN 1105  |    |   |
| Qy  | 241 QDLVGWQAPPGARSILTPCTCGSSDLYLVTRHADVIPVRRGDSRGSLLSPRPVSYLKGS 300  |    |   |

|  |   |  |  |
|--|---|--|--|
| DR   | Pfam; PF01538; HCV_NS2; 1.  | DR   | GO; GO:0016021; C:integral to membrane; IEA.                                     |
| DR   | Pfam; PF02907; HCV_NS3; 1.  | DR   | GO; GO:0019028; C:viral capsid; IEA.   |
| DR   | Pfam; PF01006; HCV_NS4a; 1.   | DR   | GO; GO:0019031; C:viral envelope; IEA.   |
| DR   | Pfam; PF01001; HCV_NS4b; 1.   | DR   | GO; GO:0005524; F:AMP binding; IEA.  |
| DR   | Pfam; PF01506; HCV_NS5a; 1.   | DR   | GO; GO:0008026; F:AMP dependent helicase activity; IEA.                          |
| DR   | Pfam; PF00271; helicase_C; 1.   | DR   | GO; GO:0005489; F:electron transporter activity; IEA.                            |
| DR   | Pfam; PF00998; viral_RdRp; 1.   | DR   | GO; GO:0016787; F:hydrolase activity; IEA.                                       |
| DR   | SMART; SM00487; DEXDC; 1.   | DR   | GO; GO:0003723; F:RNA binding; IEA.  |
| DR   | PROSITE; PS00190; CYTOCHROME_C; 1.  | DR   | GO; GO:003958; F:RNA-directed RNA polymerase activity; IEA.                      |
| KW   | ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;  | DR   | GO; GO:0008236; F:serine-type peptidase activity; IEA.                           |
| KW   | Hydrolase; Nonstructural protein; Polprotein; Transferase; Transmembrane.   | DR   | GO; GO:0005198; F:structural molecule activity; IEA.                             |
| KW   | RNA-directed RNA polymerase; Transferase; Transmembrane.  | DR   | GO; GO:0016740; F:transferase activity; IEA.                                     |
| SEQUENCE   | 3010 AA; 326801 MW; 9FEEBD1B9BB7A4B CRC64;  | DR   | GO; GO:006118; P:electrolysis and peptidolysis; IEA.                             |
| QY   | 1 VRGGDAAILLTCAVHPPELFDITKLLAIFGPMVLQAGITKPYFVRAQGLRACHMLV 60   | DR   | GO; GO:0006508; P:transcription; IEA.  |
| QY   | 96.6%; Score 1718; DB 12; Length 3010;  | DR   | GO; GO:0006350; P:viral genome replication; IEA.                                 |
| Best Local Similarity 95.6%; Pred. No. 1.2e-137; Mismatches 8; Indels 0; Gaps 0; | Matches 325; Conservative 7; Mismatches 8; Indels 0; Gaps 0;  | DR   | GO; GO:0019079; P:viral transformation; IEA.                                     |
| Db   | 866 VRGGDAAILLTCAVHPPELFDITKLLAIFGPMVLQAGITKPYFVRAQGLRACHMLV 925  | DR   | InterPro; IPR009003; CYS_Ser_trypsin.  |
| QY   | 61 RKAAGHYYVQAFMVKLAALGTGTYVVDHLPLQDWAHGLRDLAVEPVFSNEVKIT 120   | DR   | InterPro; IPR00345; CYC_heme_BS.   |
| Db   | 926 RKVAGHYVQVMAVLKLAALGTGTYVVDHLPLQDWAHGLRDLAVEPVFSNEVKIT 985  | DR   | InterPro; IPR001410; DEAD.   |
| QY   | 181 SITGRDKNQVEGEVQVSTATOSFLATCINGVCMVYHGSKTLAGPKGPITQMYNTD 240   | DR   | InterPro; IPR002521; HCV core.   |
| Db   | 1046 SITGRDKNQVEGEVQVSTATOSFLATCINGVCMVYHGSKTLAGPKGPITQMYNTD 1105   | DR   | InterPro; IPR002868; HCV capsid.   |
| QY   | 121 WGADTAACGDPISGLPVSAARRGRELILGPGDNLFGQGRLLAPITAYSSQTGIGCIT 180   | DR   | InterPro; IPR002165; HCV_RdRp.   |
| Db   | 986 WGADTAACGDPISGLPVSAARRGRELILGPGDNLFGQGRLLAPITAYSSQTGIGCIT 1045  | DR   | InterPro; IPR001650; Helicase_C.   |
| QY   | 241 QDLVGWQAPPGARSMTCTGGSDLYLVTRHADVPIVRRGSRSLSPPRVSYLKSS 300   | DR   | InterPro; IPR002531; HCV_NS1.  |
| Db   | 1106 LDLGWQAPPGARSMTCTGGSDLYLVTRHADVPIVRRGSRSLSPPRVSYLKSS 1165  | DR   | InterPro; IPR000745; HCV_NS4a.   |
| QY   | 301 GGPLICPSGHAVGFLRAAVCTRGYAKAVDPIPVSMEITMR 341  | DR   | InterPro; IPR00794; RNA_Dol_PSVir.   |
| Db   | 1166 GGPULCPGSHAVGFLRAAVCTRGYAKAVDPIPVSMEITMR 1206  | DR   | InterPro; IPR00794; RNA_Dol_PSVir.   |
| RESULT 7   |   | DR   | InterPro; IPR001490; HCV_NS4b.   |
| Q9J3P4   | PRELIMINARY; PRT; 3008 AA.  | DR   | InterPro; IPR002868; HCV capsid.   |
| ID   | Q9J3P4  | DR   | Pfam; PF01543; HCV core; 1.  |
| AC   | 00J3P4  | DR   | InterPro; IPR00251; HCV_NS1.   |
| DT   | 01-OCT-2000 (TREMBREL_15, Created)  | DR   | InterPro; IPR004109; Peptidase_C29.  |
| DT   | 01-OCT-2000 (TREMBREL_15, Last sequence update)   | DR   | InterPro; IPR00705; RNA_pol_D5_ps.   |
| DT   | 01-OCT-2003 (TREMBREL_25, Last annotation update)   | DR   | InterPro; IPR00794; RNA_Dol_PSVir.   |
| DE   | Genome polyprotein.   | DR   | Pfam; PF01006; HCV_NS3; 1.   |
| GN   | MD34.   | DR   | Pfam; PF01006; HCV_NS4; 1.   |
| OS   | Hepatitis C virus.  | DR   | Pfam; PF01001; HCV_NS4b; 1.  |
| OC   | Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;   | DR   | Pfam; PF01506; HCV_NS5a; 1.  |
| OC   | Hepatitisvirus.   | DR   | Pfam; PF00271; helicase_C; 1.  |
| NCBI_TAXID=11103;  |   | DR   | Pfam; PF00998; viral_RdRp; 1.  |
| RN   | [1]   | DR   | SMART; SM00487; DEXDC; 1.  |
| RP   | SEQUENCE FROM N.A.  | DR   | PROSITE; PS00190; CYTOCHROME_C; 1.   |
| RC   | Nagayama K., Kuroski M., Enomoto N., Miyasaka Y., Marumo F., Sato C.;   | DR   | ATP-binding; Coat protein; Envelope protein; Glycoprotein; Helicase;             |
| RT   | "Characteristics of hepatitis C viral genome associated with disease progression.";   | DR   | Hydrolase; Nonstructural protein; Polprotein; Transferase; Transmembrane.        |
| RT   | Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.   | DR   | RNA-directed RNA polymerase; Transferase; Transmembrane.                         |
| RL   | - SUBUNIT. THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND M RNA (BY SIMILARITY). | DR   | SEQUENCE 3008 AA; 326834 MW; 99AE09E14C3109F4 CRC64;                             |
| CC   | EMBL; AF008024; ABF1205.1; -.   | DR   | Query Match 96.5%; Score 1715; DB 12; Length 3008;                               |
| CC   | PIR; A61196; A61195.  | DR   | Best Local Similarity 95.6%; Pred. No. 2.1e-137; Mismatches 9; Indels 0; Gaps 0; |
| CC   | DR  | Matches 326; Conservative 6; Mismatches 9; Indels 0; Gaps 0; |  |
| DR   | PS0246; P00266.   | DR   | QY 1 VRGGDAAILLTCAVHPPELFDITKLLAIFGPMVLQAGITKPYFVRAQGLRACHMLV 60                 |
| DR   | PIR; PS0329; PS0329.  | DR   | QY 61 RKAAGHYYVQAFMVKLAALGTGTYVVDHLPLQDWAHGLRDLAVEPVFSNEVKIT 120                 |
| DR   | HSSP; P2663; 1JXP.  | DR   | QY 924 RKVAGHYVQVMAVLKLAALGTGTYVVDHLPLQDWAHGLRDLAVEPVFSNEVKIT 983                |
| DR   |   | DR   | QY 121 WGADTAACGDPISGLPVSAARRGRELILGPGDNLFGQGRLLAPITAYSSQTGIGCIT 180             |
| DR   |   | DR   | QY 984 WGADTAACGDPISGLPVSAARRGRELILGPGDNLFGQGRLLAPITAYSSQTGIGCIT 1043            |
| DR   |   | DR   | QY 181 SITGRDKNQVEGEVQVSTATOSFLATCINGVCMVYHGSKTLAGPKGPITQMYNTD 240               |
| DR   |   | DR   | QY 1044 SITGRDKNQVEGEVQVSTATOSFLATCINGVCMVYHGSKTLAGPKGPITQMYNTD 1105             |
| DR   |   | DR   | QY 241 QDLVGWQAPPGARSMTCTGGSDLYLVTRHADVPIVRRGSRSLSPPRVSYLKSS 300                 |

|    |   |   |   |
|----|---|---|---|
| Db | 1104  | QDINGWQAPPGARSLLTPTCTGSSSLYLVTRHAIIVIPVRRGDSRGSLSLSPRVSYLKGS 1163 | DR Pfam: PF01006; HCV_NS4a; 1.  |
| Qy | 301   | GGLPLCPGHAVGIFRAVCIRGVAKAVDVFVPEVESMETMR 341                      | DR Pfam; PF01506; HCV_NS5a; 1.  |
| Db | 1164  | GGLPLCPGHAVGIFRAVCIRGVAKAVDVFVPEVESMETMR 1204                     | DR Pfam; PF00271; helicase_C; 1.  |
|    |   |   | DR Pfam; PF0998; viral_RBP; 1.  |
|    |   |   | DR PRODOM; P0186062; HCV_NS1; 1.  |
|    |   |   | DR SMART; SM0487; DEADC; 1.   |
|    |   |   | DR PROSITE; PS00190; CYTOCHROME_C; 1.                                   |
|    |   |   | DR Coat_protein; Envelope_protein; Glycoprotein; Nonstructural_protein; |
|    |   |   | KM Polyprotein; RNA-directed_RNA_polymerase; Transmembrane.             |
| ID | Q9J3H3  | PRELIMINARY; PRT; 3010 AA.  | SQ SEQUENCE   |
| AC | Q9J3H3;   |   |   |
| DT | 01-OCT-2000 (TREMBrel. 15, Created)   |   |   |
| DT | 01-OCT-2003 (TREMBrel. 25, Last annotation update)  |   |   |
| DE | Genome polyprotein.   |   |   |
| OS | Hepatitis C virus.  |   |   |
| OC | Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;   |   |   |
| OC | Reoviridae.   |   |   |
| OX | NCBI_TaxID=11103;   |   |   |
| RN | {1}   |   |   |
| RC | SEQUENCE FROM N.A.  |   |   |
| RA | STRAIN=MD19;  |   |   |
| RA | Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.,   |   |   |
| RT | "Characteristics of hepatitis C viral genome associated with disease progression.",   |   |   |
| RL | Submission (Nov-1999) to the EMBL/GenBank/DDBJ databases.   |   |   |
| CC | - SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: |   |   |
| CC | PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND mRNA. (BY SIMILARITY).                             |   |   |
| CC | EMBL: AF207760; AAC65950.1; -.  |   |   |
| DR | PIR: A61196; A61196.  |   |   |
| DR | P50322;   |   |   |
| DR | HSSP; P26663; IJXP.   |   |   |
| DR | GO: GO_0019028; C:integral membrane; IEA.   |   |   |
| DR | GO; GO_000524; F-ATP binding; IEA.  |   |   |
| DR | GO; GO_0008026; F-ATP dependent helicase activity; IEA.   |   |   |
| DR | GO; GO_0003723; F:electron transporter activity; IEA.   |   |   |
| DR | GO; GO_0003668; F-RNA-directed RNA polymerase activity; IEA.  |   |   |
| DR | GO; GO_0008226; F-serrine-type peptidase activity; IEA.   |   |   |
| DR | GO; GO_0005198; F-structural molecule activity; IEA.  |   |   |
| DR | GO; GO_0016749; F-transferase activity; IEA.  |   |   |
| DR | GO; GO_0006118; Pelecan trapdoor; IEA.  |   |   |
| DR | GO; GO_0006508; P:proteinolysis and peptidolysis; IEA.  |   |   |
| DR | GO; GO_0006550; P:transcription; IEA.   |   |   |
| DR | GO; GO_0019079; P:viral genome replication; IEA.  |   |   |
| DR | GO; GO_0019057; P:viral transformation; IEA.  |   |   |
| DR | InterPro; IPR00903; Cys_Ser_trypsin.  |   |   |
| DR | InterPro; IPR00345; CytC_home_3S.   |   |   |
| DR | InterPro; IPR001410; DEAD.  |   |   |
| DR | InterPro; IPR002522; HCV_capsid.  |   |   |
| DR | InterPro; IPR00221; HCV core.   |   |   |
| DR | InterPro; IPR002519; HCV_env.   |   |   |
| DR | InterPro; IPR002521; HCV_NS1.   |   |   |
| DR | InterPro; IPR002518; HCV_NS2.   |   |   |
| DR | InterPro; IPR00745; HCV_NS3.  |   |   |
| DR | InterPro; IPR001450; HCV_NS4B.  |   |   |
| DR | InterPro; IPR02868; HCV_NS5A.   |   |   |
| DR | InterPro; IPR02156; HCV_RBP.  |   |   |
| DR | InterPro; IPR001650; Helicase_C.  |   |   |
| DR | InterPro; IPR004109; Peptidase_C29.   |   |   |
| DR | InterPro; IPR00705; RNA_pol_DS_Ps.  |   |   |
| DR | InterPro; IPR007094; RNA_Pol_PsVir.   |   |   |
| DR | Pfam; PF011543; HCV_capsid; 1.  |   |   |
| DR | Pfam; PF01542; HCV_core; 1.   |   |   |
| DR | Pfam; PF01539; HCV_env; 1.  |   |   |
| DR | Pfam; PF01560; HCV_NS1; 1.  |   |   |
| DR | Pfam; PF01538; HCV_NS2; 1.  |   |   |
| DR | Pfam; PF02907; HCV_NS3; 1.  |   |   |
|    | GO; GO_0016021; C:integral membrane; IEA.   |   |   |

RESULT 8

Q9J3H3 PRELIMINARY; PRT; 3010 AA.

AC Q9J3H3;

DT 01-OCT-2000 (TREMBrel. 15, Last sequence update)

DE Genome polyprotein.

OS Hepatitis C virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;

OC Reoviridae.

OX NCBI\_TaxID=11103;

RN {1}

RC SEQUENCE FROM N.A.

RA Kurosaki M., Enomoto N., Miyasaka Y., Marumo F., Sato C.,

RT "Characteristics of hepatitis C viral genome associated with disease progression.",

RL Submission (Nov-1999) to the EMBL/GenBank/DDBJ databases.

CC - SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:

CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF PROTEIN C AND mRNA. (BY SIMILARITY).

CC EMBL: AF207760; AAC65950.1; -.

DR PIR: A61196; A61196.

DR PIR; PQ0804; PQ0804.

DR PIR; PS0329; PS0329.

DR HSSP; P26663; IJXP.

DR MEROPS; S28\_002; -.

DR GOPO; S28\_002; -.

DR GO; GO\_0016021; C:integral membrane; IEA.

DR GO; GO:0019028; C:viral capsid; IEA.  
 DR GO; GO:0019031; C:viral envelope; IEA.  
 DR GO; GO:0005524; F:ATP dependent helicase activity; IEA.  
 DR GO; GO:0008026; F:ATP binding; IEA.  
 DR GO; GO:0005489; F:electron transporter activity; IEA.  
 DR GO; GO:0003723; F:RNA binding; IEA.  
 DR GO; GO:0003968; F:RNA polymerase activity; IEA.  
 DR GO; GO:0008235; F:serine-type peptidase activity; IEA.  
 DR GO; GO:0005198; F:structural molecule activity; IEA.  
 DR GO; GO:0016740; F:transferase activity; IEA.  
 DR GO; GO:0006118; F:electron transport; IEA.  
 DR GO; GO:0006508; F:peptidolysis and peptidolysis; IEA.  
 DR GO; GO:0019079; F:viral genome replication; IEA.  
 DR GO; GO:0019087; F:viral transformation; IEA.  
 DR InterPro; IPR009003; Cys\_Ser\_trypsin.  
 DR InterPro; IPR000345; CytC\_heme\_BS.  
 DR InterPro; IPR01410; DEAD.  
 DR InterPro; IPR02522; HCV\_capsid.  
 DR InterPro; IPR00231; HCV\_core.  
 DR InterPro; IPR02531; HCV\_env.  
 DR InterPro; IPR02518; HCV\_NS1.  
 DR InterPro; IPR000745; HCV\_NS2.  
 DR InterPro; IPR01490; HCV\_NS3b.  
 DR InterPro; IPR02868; HCV\_NS5a.  
 DR InterPro; IPR02166; HCV\_RdRp.  
 DR InterPro; IPR001650; Helicase\_C.  
 DR InterPro; IPR04109; Peptidase\_C9.  
 DR InterPro; IPR007035; RNA\_pol\_DS\_PS.  
 DR InterPro; IPR007094; RNA\_pol\_PSVir.  
 DR Pfam; PR011543; HCV\_capsid; 1.  
 DR Pfam; PR012166; HCV\_core; 1.  
 DR Pfam; PR001650; Helicase\_C.  
 DR Pfam; PR01580; HCV\_NS1; 1.  
 DR Pfam; PR01538; HCV\_NS2; 1.  
 DR Pfam; PR02907; HCV\_NS3; 1.  
 DR Pfam; PR01106; HCV\_NS4a; 1.  
 DR Pfam; PR01542; HCV\_core; 1.  
 DR Pfam; PR01539; HCV\_env; 1.  
 DR Pfam; PR00271; helicase\_C; 1.  
 DR Pfam; PR00998; viral\_RdRp; 1.  
 DR Prodrom; PDI18662; HCV\_NS1; 1.  
 DR SMART; SM00487; DEKDC\_1.  
 DR PROSITE; PS00190; CYTOCHROME\_C; 1.  
 DR KW Coat protein; Envelope protein; Glycoprotein; Nonstructural protein; Polyproteins; RNA-directed RNA polymerase; Transferase; Transmembrane; Sequence; 301 AA; 326909 MW; 550562EBODAC519 CRC64;

Query Match 96.5%: Score 1715; DB 12; Length 301; Best Local Similarity 95.3%; Pred. No. 2.1e-137; Matches 325; Conservative 8; Mismatches 8; Indels 0; Gaps 0; SQ

1 VRGERDAILLLTCVHPELFDITKLIAIFGLMLQDAGITKPYFVRAGLIRACMV  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
 866 VRRGRDAILLLTCVHPELFDITKLIAIFGPMIVLQAGITRVPYFVPAQSLIRACMV  
 925 61 RKAAGHYYQMAFMKLAIGTGYDHLPQLQDMAHAGLIRDIAVAVPVSFSDMEVKIT  
 ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||  
 926 RKAAGHYYQMAFMKLAIGTGYDHLPQLQDMAHAGLIRDIAVAVPVSFSDMEVKIT 985  
 DR GO; GO:0005489; F:electron transporter activity; IEA.  
 DR InterPro; IPR02531; HCV\_env.  
 DR InterPro; IPR002518; HCV\_NS2.  
 DR InterPro; IPR000745; HCV\_NS4a.  
 DR InterPro; IPR001490; HCV\_NS4b.  
 DR InterPro; IPR002868; HCV\_NS5a.  
 DR InterPro; IPR002166; HCV\_RdRp.  
 DR InterPro; IPR04109; Peptidase\_C9.  
 DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro; IPR007094; RNA\_pol\_PSVir.  
 DR Pfam; PR01543; HCV\_capsid; 1.  
 DR Pfam; PR01542; HCV\_core; 1.  
 DR Pfam; PR01539; HCV\_env; 1.  
 DR Pfam; PR01560; HCV\_NS1; 1.  
 DR Pfam; PR01538; HCV\_NS2; 1.  
 DR Pfam; PR00998; viral\_RdRp; 1.  
 DR Prodrom; PDI18662; HCV\_NS1; 1.  
 DR SMART; SM00487; DEKDC\_1.  
 DR PROSITE; PS00190; CYTOCHROME\_C; 1.  
 DR KW Polyprotein.

Qy 301 GGPLICPSGHAVGIRAVCTRGVAKAVDPIPVSMEETMR 341  
 Db 1166 GGPLUCPSGHAVGIRAVCTRGVAKAVDPIPVSMEETMR 1206

RESULT 10  
 ID Q8073 PRELIMINARY; PRT; 3010 AA.  
 AC Q8073; DT 01-JUN-2003 (TREMBirel. 24, Created)  
 DT 01-OCT-2003 (TREMBirel. 25, Last sequence update)  
 DE Polypeptide.  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 Hepacivirus.  
 OC OK NCBI\_TaxID=11103;  
 RN [1] IPID  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MILE;  
 RA Kishino H., Sugiyama K., Hijikata M., Kato N., Takahashi H., Noshi T.,  
 Nio Y., Hosaka M., Miyazaki Y., Shimochono K.; Infected with the  
 hepatitis C virus.", Bioclim. Biophys. Res. Commun. 293:93-99(2002).  
 RL EMBL; AB002299; BA54896.1; -.  
 DR GO; GO:0019028; C:viral capsid; IEA.  
 DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
 DR GO; GO:0008235; F:serine-type peptidase activity; IEA.  
 DR GO; GO:0005524; F:ATP dependent helicase activity; IEA.  
 DR GO; GO:0008026; F:structural molecule activity; IEA.  
 DR GO; GO:0005489; F:electron transporter activity; IEA.  
 DR GO; GO:0003723; F:peptidolysis and peptidolysis; IEA.  
 DR GO; GO:0006350; F:transcription; IEA.  
 DR GO; GO:0019079; F:viral genome replication; IEA.  
 DR GO; GO:0019087; F:viral transformation; IEA.  
 DR InterPro; IPR009003; Cys\_Ser\_trypsin.  
 DR InterPro; IPR00345; CytC\_heme\_BS.  
 DR InterPro; IPR00222; HCV\_capsid.  
 DR InterPro; IPR00251; HCV\_core.  
 DR InterPro; IPR002521; HCV\_env.  
 DR InterPro; IPR02531; HCV\_NS1.  
 DR InterPro; IPR00410; Peptidase\_C9.  
 DR InterPro; IPR007095; RNA\_pol\_DS\_PS.  
 DR InterPro; IPR007094; RNA\_pol\_PSVir.  
 DR Pfam; PR01543; HCV\_capsid; 1.  
 DR Pfam; PR01542; HCV\_core; 1.  
 DR Pfam; PR01539; HCV\_env; 1.  
 DR Pfam; PR01560; HCV\_NS1; 1.  
 DR Pfam; PR01538; HCV\_NS2; 1.  
 DR Pfam; PR00998; viral\_RdRp; 1.  
 DR Prodrom; PDI18662; HCV\_NS1; 1.  
 DR SMART; SM00487; DEKDC\_1.  
 DR PROSITE; PS00190; CYTOCHROME\_C; 1.  
 DR KW Polyprotein.

|                       |   |                                 |                               |  |      |
|-----------------------|---|---------------------------------|-------------------------------|--|------|
| Qy                    | 1 VRGERGDAILITCAVPHLIFDITKLLIAIFGPMLVQAGIKTPYVFRAGLIRACMV                       | 60                              | DR                            | GO; GO:0016740; F-Transferase activity; IEA.                         |      |
| :                     | : | 955                             | DR                            | GO; GO:0005508; Proteolysis and peptidolysis; IEA.                   |      |
| Best Local Similarity | 95.5%;  | Score 1715; DB 12; Length 3010; | DR                            | GO; GO:0006350; PTtranscription; IEA.                                |      |
| Matches               | 326;  | Conservative                    | DR                            | GO; GO:0019079; P-viral genome replication; IEA.                     |      |
| Dr                    | 6; Mismatches   | 9; Indels 0; Gaps 0;            | DR                            | GO; GO:0019087; P-viral transformation; IEA.                         |      |
| Qy                    | 1 VRGERGDAILITCAVPHLIFDITKLLIAIFGPMLVQAGIKTPYVFRAGLIRACMV                       | 60                              | DR                            | InterPro; IPR009003; Cys Ser_trypsin.                                |      |
| :                     | : | 925                             | DR                            | InterPro; IPR001410; D2AD.   |      |
| B66                   | IRGGDAITLMCAHPLPELFIDITKLLIAIFGPMLVQAGIKTPYVFRAGLIRACMV                         | 925                             | DR                            | InterPro; IPR002522; HCV_capsid.                                     |      |
| DR                    | WGADTAACGDIILGLJPVSARRGRETTTGPADSQCGWRLAPITYSQOTRGIGCIT                         | 1045                            | DR                            | InterPro; IPR002521; HCV_core.                                       |      |
| Qy                    | 61 RKAAGHYVQAFMVKLAALTGTGVYDHLPQLDWAHAGLIRDAVAPVPSDMETKLT                       | 120                             | DR                            | InterPro; IPR002319; HCV_env.  |      |
| 986                   | RKAAGHYVQAFMVKLAALTGTGVYDHLPQLDWAHAGLIRDAVAPVPSDMETKLT                          | 120                             | DR                            | InterPro; IPR002311; HCV_NS1.  |      |
| DR                    | 926 RKAAGHYVQAFMVKLAALTGTGVYDHLPQLDWAHAGLIRDAVAPVPSDMETKLT                      | 985                             | DR                            | InterPro; IPR002518; HCV_NS4a.                                       |      |
| DR                    | 181 SLTERBDKNQVEGEVQVSVQVSTATOSFLATCINGVQWVPHAGSKLAGPQPKITQMTND                 | 240                             | DR                            | InterPro; IPR000745; RNA_poli_Ds_ps.                                 |      |
| DR                    | 1046 SLTGRDKNQVEGEVQVSVQVSTATOSFLATCINGVQWVPHAGSKLAGPQPKITQMTND                 | 1105                            | DR                            | InterPro; IPR007094; RNA_poli_Psvir.                                 |      |
| Db                    | 241 QDLVGWQAPGARSMPCTCGSSDLYLVTRADVTPVRRGDSRGSLSPPRVSYLKGS                      | 300                             | DR                            | InterPro; IPR002166; HCV_Ramp.                                       |      |
| Qy                    | 301 GGPILCPSPGHAVGIFRAVCTRGVAKAVDPFIPVESMETMR                                   | 341                             | DR                            | InterPro; IPR001650; Helicase_C.                                     |      |
| Qy                    | 1166 GGPILCPSPGHAVGIFRAVCTRGVAKAVDPFIPVESMETMR                                  | 1206                            | DR                            | InterPro; IPR004109; peptidase_C29.                                  |      |
| DR                    | PFam; PF00938; HCV_NS2; 1.  | DR                              | Pfam; PF01538; HCV_NS2; 1.    |  |      |
| DR                    | PFam; PF02907; HCV_NS3; 1.  | DR                              | Pfam; PF01006; HCV_NS3; 1.    |  |      |
| DR                    | PFam; PF01542; HCV_core; 1.   | DR                              | Pfam; PF01542; HCV_core; 1.   |  |      |
| DR                    | PFam; PF01539; HCV_env; 1.  | DR                              | Pfam; PF01539; HCV_env; 1.    |  |      |
| DR                    | PFam; PF01560; HCV_NS1; 1.  | DR                              | Pfam; PF01560; HCV_NS1; 1.    |  |      |
| DR                    | PFam; PF01538; HCV_NS2; 1.  | DR                              | Pfam; PF01538; HCV_NS2; 1.    |  |      |
| DR                    | PFam; PF01006; HCV_NS3; 1.  | DR                              | Pfam; PF01006; HCV_NS3; 1.    |  |      |
| DR                    | PFam; PF01001; HCV_NtAb; 1.   | DR                              | Pfam; PF01001; HCV_NtAb; 1.   |  |      |
| DR                    | PFam; PF01506; HCV_NS3a; 1.   | DR                              | Pfam; PF01506; HCV_NS3a; 1.   |  |      |
| DR                    | PFam; PF00271; helicase_C; 1.   | DR                              | Pfam; PF00271; helicase_C; 1. |  |      |
| DR                    | PFam; PF00938; viral_RdRp; 1.   | DR                              | Pfam; PF00938; viral_RdRp; 1. |  |      |
| DR                    | PRODom; PD186062; HCV_NS1; 1.   | DR                              | PRODom; PD186062; HCV_NS1; 1. |  |      |
| DR                    | SMART; SM00487; DEXD_C; 1.  | DR                              | SMART; SM00487; DEXD_C; 1.    |  |      |
| KW                    | Coat protein; Envelope protein; Glycoprotein; Nonstructural protein;            | DR                              | KW                            | Coat protein; Envelope protein; Glycoprotein; Nonstructural protein; |      |
| KW                    | Polyprotein; RNA-directed RNA polymerase; Transfase; Transmembrane.             | DR                              | KW                            | Polyprotein; RNA-directed RNA polymerase; Transfase; Transmembrane.  |      |
| FT                    | CHAIN 1 191 CORE PROTEIN.   | DR                              | FT                            | CHAIN 1 191 CORE PROTEIN.  |      |
| FT                    | CHAIN 192 383 EL1.  | DR                              | FT                            | CHAIN 192 383 EL1.   |      |
| FT                    | CHAIN 384 809 EL2.  | DR                              | FT                            | CHAIN 384 809 EL2.   |      |
| FT                    | CHAIN 810 1025 NS2.   | DR                              | FT                            | CHAIN 810 1025 NS2.  |      |
| FT                    | CHAIN 1027 1657 NS3.  | DR                              | FT                            | CHAIN 1027 1657 NS3.   |      |
| FT                    | CHAIN 1658 1711 NS4A.   | DR                              | FT                            | CHAIN 1658 1711 NS4A.  |      |
| FT                    | CHAIN 1712 1972 NS4B.   | DR                              | FT                            | CHAIN 1712 1972 NS4B.  |      |
| FT                    | CHAIN 1973 2419 NS5A.   | DR                              | FT                            | CHAIN 1973 2419 NS5A.  |      |
| FT                    | CHAIN 2420 3010 NS5B.   | DR                              | FT                            | CHAIN 2420 3010 NS5B.  |      |
| SQ                    | SEQUENCE 3010 AA; 327332 MW; 5F81505783F8FPB8 CRC64;                            | DR                              | SQ                            | SEQUENCE 3010 AA; 327332 MW; 5F81505783F8FPB8 CRC64;                 |      |
| Query Match           | 96.4%; Score 1714; DB 12; Length 3010;  | DR                              | Query Match                   | 96.4%; Score 1714; DB 12; Length 3010;                               |      |
| Best Local Similarity | 94.7%; Pred. No. 2 6e-137;  | DR                              | Best Local Similarity         | 94.7%; Pred. No. 2 6e-137;   |      |
| Matches               | 323; Conservative 12; Mismatches 6; Indels 0; Gaps 0;                           | DR                              | Matches                       | 323; Conservative 12; Mismatches 6; Indels 0; Gaps 0;                |      |
| Qy                    | 1 VRGERGDAILITCAVPHLIFDITKLLIAIFGPMLVQAGIKTPYVFRAGLIRACMV                       | 60                              | Qy                            | 1 VRGERGDAILITCAVPHLIFDITKLLIAIFGPMLVQAGIKTPYVFRAGLIRACMV            | 60   |
| DR                    | 866 VRGERGDAILITCAVPHLIFDITKLLIAIFGPMLVQAGIKTPYVFRAGLIRACMV                     | 925                             | DR                            | 866 VRGERGDAILITCAVPHLIFDITKLLIAIFGPMLVQAGIKTPYVFRAGLIRACMV          | 925  |
| Qy                    | 61 RKAAGHYVQAFMVKLAALTGTGVYDHLPQLDWAHAGLIRDAVAPVPSDMETKLT                       | 120                             | Qy                            | 61 RKAAGHYVQAFMVKLAALTGTGVYDHLPQLDWAHAGLIRDAVAPVPSDMETKLT            | 120  |
| Db                    | 926 RKAAGHYVQAFMVKLAALTGTGVYDHLPQLDWAHAGLIRDAVAPVPSDMETKLT                      | 985                             | Db                            | 926 RKAAGHYVQAFMVKLAALTGTGVYDHLPQLDWAHAGLIRDAVAPVPSDMETKLT           | 985  |
| DR                    | 121 WGADTAACGDIISGLPYSARRGRETTTGPADSQCGWRLAPITYSQOTRGIGCIT                      | 180                             | DR                            | 121 WGADTAACGDIISGLPYSARRGRETTTGPADSQCGWRLAPITYSQOTRGIGCIT           | 180  |
| DR                    | 181 SLTGRDKNQVEGEVQVSVQVSTATOSFLATCINGVQWVPHAGSKLAGPQPKITQMTND                  | 240                             | DR                            | 181 SLTGRDKNQVEGEVQVSVQVSTATOSFLATCINGVQWVPHAGSKLAGPQPKITQMTND       | 240  |
| Db                    | 1046 SLTGRDKNQVEGEVQVSVQVSTATOSFLATCINGVQWVPHAGSKLAGPQPKITQMTND                 | 1105                            | Db                            | 1046 SLTGRDKNQVEGEVQVSVQVSTATOSFLATCINGVQWVPHAGSKLAGPQPKITQMTND      | 1105 |
| Qy                    | 241 QDLVGWQAPGARSMPCTCGSSDLYLVTRADVTPVRRGDSRGSLSPPRVSYLKGS                      | 300                             | Qy                            | 241 QDLVGWQAPGARSMPCTCGSSDLYLVTRADVTPVRRGDSRGSLSPPRVSYLKGS           | 300  |
| DR                    | 301 GGPILCPSPGHAVGIFRAVCTRGVAKAVDPFIPVESMETMR                                   | 341                             | DR                            | 301 GGPILCPSPGHAVGIFRAVCTRGVAKAVDPFIPVESMETMR                        | 341  |
| DR                    | 1166 GGPILCPSPGHAVGIFRAVCTRGVAKAVDPFIPVESMETMR                                  | 1206                            | DR                            | 1166 GGPILCPSPGHAVGIFRAVCTRGVAKAVDPFIPVESMETMR                       | 1206 |

**RESULT 12**

ID Q81755 PRELIMINARY; PRT; 1186 AA.

AC 081755;

DT 01-NOV-1996 (TREMBREL\_01, Last sequence update)

DT 01-OCT-2003 (TREMBREL\_25, Last annotation update)

DE Polyprotein (Fragment).

OS Hepatitis C virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepacivirus.

OX NCBI\_TAXID=1103;

[1] SEQUENCE FROM N.A. RP

RX MEDLINE=9108950; PubMed=2175903; Kato N., Hijikata M., Ootsuyama Y., Nakagawa M., Ohkoshi S., Sugimura T., Shimotohno K.; "Molecular cloning of the human hepatitis C virus genome from Japanese patients with non-A, non-B hepatitis"; Proc. Natl. Acad. Sci. U.S.A. 87:9524-9528(1990).

RN [2] SEQUENCE FROM N.A. RP

RX MEDLINE=9204440; PubMed=158196; Okamoto H., Okada S., Sugiyama Y., Kurai K., Iizuka H., Machida A., Miyakawa Y., Mayumi M.; "Nucleotide sequences of the genomic RNA of hepatitis C virus isolated from a human carrier: comparison with reported isolates for conserved and divergent regions"; J. Gen. Virol. 72:2697-2704(1991).

RN [3] SEQUENCE FROM N.A. RP

RX MEDLINE=9114098; PubMed=1847440; Takemizawa A., Mori C., Manabe S., Murakami S., Fujita J., Onishi E., Andoh T., Yoshida I., Okajima H.; "The structure and organization of the Hepatitis C virus genome isolated from human carriers"; J. Virol. 65:1105-1113(1991).

RN [4] SEQUENCE FROM N.A. RP

RX MEDLINE=9117826; PubMed=1848704; Choo Q.-L., Richman K., Han J.H., Berger K., Lee C., Dong C., Gallegos C., Coit D., Medina-Selby A., Barr P.J., Weiner A., Bradley D.W., Kuo G., Houghton M.; "Genetic organization and diversity of the hepatitis C virus genome mapping the 5' termini of viral genomic and antigenomic RNA"; Proc. Natl. Acad. Sci. U.S.A. 88:2451-2455(1991).

RN [5] SEQUENCE FROM N.A. RP

RX MEDLINE=9223006; PubMed=1314449; Chen P., Lin M., Tai K., Liu P., Lin C., Chen D.; "The Taiwanese hepatitis C virus genome: Sequence determination and mapping the 5' termini of viral genomic and antigenomic RNA"; Virology 188:102-113(1992).

RN [6] SEQUENCE FROM N.A. RP

RX MEDLINE=9223032; PubMed=1314459; Okamoto H., Kurai K., Okada S., Yamamoto K., Iizuka H., Tanaka T., Fukuda S., Tsuda F., Mishiro S.; "Full-length sequence of a hepatitis C virus genome having poor homology to reported isolates: Comparative study of four distinct genotypes"; Virology 188:331-341(1992).

RN [7] SEQUENCE FROM N.A. RP

RX MEDLINE=93323208; PubMed=8392605; Hijikata M., Mizushima H., Almogi T., Mori S., Kakuchi N., Kato N., Tanaka T., Kimura K., Shimotohno K.; "Two distinct proteinase activities required for the processing of a putative nonstructural precursor protein of hepatitis C virus"; J. Virol. 67:4665-4675(1993).

RN [8] SEQUENCE FROM N.A. RP

RN [9] SEQUENCE FROM N.A. RP

RN [10] SEQUENCE FROM N.A. RP

RN [11] SEQUENCE FROM N.A. RP

**RESULT 13**

ID Q81817 PRELIMINARY; PRT; 2284 AA.

AC 081817;

DT 01-NOV-1996 (TREMBREL\_01, Created)

DT 01-NOV-1996 (TREMBREL\_01, Last sequence update)

DT 01-OCT-2003 (TREMBREL\_25, Last annotation update)

DR GO; GO:0005524; F-ATP binding; IEA.

DR GO; GO:0008025; F-ATP dependent helicase activity; IEA.

DR GO; GO:0016787; F:Hydrolase activity; IEA.

DR GO; GO:0003676; F:nucleic acid binding; IEA.

DR GO; GO:0008236; F:serine-type peptidase activity; IEA.

DR GO; GO:006508; F:proteolysis and peptidolysis; IEA.

DR GO; GO:0019087; F:viral transformation; IEA.

DR InterPro; IPRO09003; CYS Ser\_trypsin.

DR InterPro; IPRO01410; DEAD.

DR InterPro; IPRO02518; HCV\_NS2.

DR InterPro; IPRO00745; HCV\_NS4a.

DR InterPro; IPRO01490; HCV\_NS4b.

DR InterPro; IPRO01160; Helicase\_C.

DR InterPro; IPRO04109; Peptidase\_C29.

DR Pfam; PF01538; HCV\_NS2; 1.

DR Pfam; PF02907; HCV\_NS3; 1.

DR Pfam; PF01006; HCV\_NS4a; 1.

DR Pfam; PF00271; Helicase\_C; 1.

DR SMART; SM00487; DEADC; 1.

RW ATP-binding; Helicase; Hydrolase.

PFT NON\_TER 1 1 SQ SEQUENCE 1186 AA; 126280 MW; 34179478BA23729A CRC54;

Query Match 96.2%; Score 1711; DB 12; Length 1186; Best Local Similarity 95.3%; Pred. No. 1.4e-137; Matches 325; Conservative 9; Mismatches 0; Indels 0; Gaps 0;

Matches

QY 1 VRGGGRLLILTCATHPELFIDIKLLIATFGPMVLAGITKPYFRAQGLIRACMV 60

Db 144 VRRGGDAAITLTCAVHPFLFDITKLLIATGPMVLQAGMTRVYFRAQGLIRACMV 203

QY 61 RKAAGHGVQNAFMKLALTGTIVYDHPLQDWAHAGRLDAVAVEPVIFSMEVKIT 120

Db 204 RKVAGHGVQNAFMKLALTGTIVYDHPLQDWAHAGRLDAVAVEPVIFSMEVKIT 263

QY 121 WGADTAACGDTISGLPVPSABRGRETLIGLPGADNPFGQWLLAPITAYSQQTTRGLGCT 180

Db 264 WGADTAACGDTISGLPVPSABRGRETLIGLPGADNPFGQWLLAPITAYSQQTTRGLGCT 323

QY 181 SLTGRRNQVEGEVQVSTATOSFLATCNGVWNTVFRGAGSKTLAGPKPITQTYTNVD 240

Db 324 SLTGRRNQVEGEVQVSTATOSFLATCNGVWNTVFRGAGSKTLAGPKPITQTYTNVD 383

QY 241 QDLVGGAPPGRSMPCTGGSDIYLVTRHADVPVRGRGDSRSLSPPRVSYKSS 300

Db 384 QDLVGGAPPGRSMPCTGGSDIYLVTRHADVPVRGRGDSRSLSPPRVSYKSS 443

QY 301 GGPLUCPSGHAVGIRRAVCTRGVAKADPPIPVSMEETMR 341

Db 444 GGPLUCPSGHAVGIRRAVCTRGVAKADPPIPVSMEETMR 484

Hijikata M.; Submitted (DEC-1993) to the EMBL/GenBank/dDBJ databases.

RJ EMLB; D11397; BA20975.1; -.

DR EMBL; D11397; BA20975.1; -.

DR PIR; A61196; A61196.

DR PIR; PS0329; PS0329.

DR PDB; 1DXP; 28-MAR-02.

DR GO; GO:0005524; F-ATP binding; IEA.

DR GO; GO:0008025; F-ATP dependent helicase activity; IEA.

DR GO; GO:0016787; F:Hydrolase activity; IEA.

DR GO; GO:0003676; F:nucleic acid binding; IEA.

DR GO; GO:0008236; F:serine-type peptidase activity; IEA.

DR GO; GO:006508; F:proteolysis and peptidolysis; IEA.

DR GO; GO:0019087; F:viral transformation; IEA.

DR InterPro; IPRO09003; CYS Ser\_trypsin.

DR InterPro; IPRO01410; DEAD.

DR InterPro; IPRO02518; HCV\_NS2.

DR InterPro; IPRO00745; HCV\_NS4a.

DR InterPro; IPRO01490; HCV\_NS4b.

DR InterPro; IPRO01160; Helicase\_C.

DR InterPro; IPRO04109; Peptidase\_C29.

DR Pfam; PF01538; HCV\_NS2; 1.

DR Pfam; PF02907; HCV\_NS3; 1.

DR Pfam; PF01006; HCV\_NS4a; 1.

DR Pfam; PF00271; Helicase\_C; 1.

DR SMART; SM00487; DEADC; 1.

RW ATP-binding; Helicase; Hydrolase.

|  |  |                    |          |            |                         |
|--|--|--------------------|----------|------------|-------------------------|
| RA   | RA   | SEQUENCE FROM N.A. | 2284 AA; | 247213 MN; | DC272A1517046337 CRC64; |
| Hijikata M., Mizushima H., Tanji Y., Konada Y., Hirowatari Y.,   | Hijikata M., Mizushima H., Tanji Y., Konada Y., Hirowatari Y.,   |                    |          |            |                         |
| Akagi T., Kimura K., Shimotohno K.; "Proteolytic processing and membrane association of putative nonstructural proteins of hepatitis C virus"; Proc. Natl. Acad. Sci. U.S.A. 90:10773-10777(1993). |  |                    |          |            |                         |
| [2]  |  |                    |          |            |                         |
| RP   | SEQUENCE FROM N.A.   |                    |          |            |                         |
| RX   | MEDLINE:9433310; PubMed=8056334;   |                    |          |            |                         |
| RA   | MEDLINE:95056078; PubMed=7956638;  |                    |          |            |                         |
| RT   | "Identification of the domain required for trans-cleavage activity of hepatitis C viral serine protease.";         |                    |          |            |                         |
| RT   | hepatitis C virus polyprotein processing: kinetics and mutagenic analysis of serine protease-dependent cleavage."; |                    |          |            |                         |
| RL   | J. Virol. 68:8418-8422(1994).  |                    |          |            |                         |
| RN   | [3]  |                    |          |            |                         |
| RP   | SEQUENCE FROM N.A.   |                    |          |            |                         |
| RX   | MEDLINE:9516583; PubMed=7853491;   |                    |          |            |                         |
| RA   | Tanji Y., Hijikata M., Satoh S., Kaneko T., Shimotohno K.;   |                    |          |            |                         |
| RT   | "Hepatitis C virus-encoded nonstructural protein NS4A has versatile functions in viral protein processing.";       |                    |          |            |                         |
| RL   | J. Virol. 69:1575-1581(1995).  |                    |          |            |                         |
| EMBL   | D16435; BAA3905.1; -.  |                    |          |            |                         |
| PR   | A61196; A61196.  |                    |          |            |                         |
| DR   | P0246; P0046.  |                    |          |            |                         |
| DR   | PIR; PS0329; FS0329.   |                    |          |            |                         |
| HSRP; P26663; IJXP.  |  |                    |          |            |                         |
| DR   | GO; GO_0019012; Cvirion; IEA.  |                    |          |            |                         |
| DR   | GO; GO_0005224; FATP binding; IEA.   |                    |          |            |                         |
| DR   | GO; GO_0008026; FATP dependent helicase activity; IEA.   |                    |          |            |                         |
| DR   | GO; GO_0003723; F-RNA binding; IEA.  |                    |          |            |                         |
| DR   | GO; GO_0003668; RNA-directed RNA polymerase activity; IEA.   |                    |          |            |                         |
| DR   | GO; GO_0008236; Feserine-type peptidase activity; IEA.   |                    |          |            |                         |
| DR   | GO; GO_0016740; Ftranferase activity; IEA.   |                    |          |            |                         |
| DR   | GO; GO_000508; P:Proteolysis and peptidolysis; IEA.  |                    |          |            |                         |
| DR   | GO; GO_0006350; Pttranscription; IEA.  |                    |          |            |                         |
| DR   | GO; GO_0019079; P:viral genome replication; IEA.   |                    |          |            |                         |
| DR   | GO; GO_0019087; P:viral transformation; IEA.   |                    |          |            |                         |
| DR   | InterPro; IPR00903; CYS Ser_trypsin.   |                    |          |            |                         |
| DR   | InterPro; IPR00110; DEAD.  |                    |          |            |                         |
| DR   | InterPro; IPR002518; HCV_NS2.  |                    |          |            |                         |
| DR   | InterPro; IPR00045; HCV_NS4a.  |                    |          |            |                         |
| DR   | InterPro; IPR00190; HCV_NS4b.  |                    |          |            |                         |
| DR   | InterPro; IPR02668; HCV_Nsa.   |                    |          |            |                         |
| DR   | InterPro; IPR02166; HCV_RdRp.  |                    |          |            |                         |
| DR   | InterPro; IPR00150; Helicase C.  |                    |          |            |                         |
| DR   | InterPro; IPR00410; Peptidase C29.   |                    |          |            |                         |
| DR   | InterPro; IPR00795; RNA_p1_D5_Ps.  |                    |          |            |                         |
| DR   | InterPro; IPR00794; RNA_pol_Psvr.  |                    |          |            |                         |
| DR   | Fram; PF01338; HCV_NS2; 1.   |                    |          |            |                         |
| DR   | Fram; PF02907; HCV_NS3; 1.   |                    |          |            |                         |
| DR   | Fram; PF0106; HCV_Nsa; 1.  |                    |          |            |                         |
| DR   | Fram; PF01001; HCV_Nsb; 1.   |                    |          |            |                         |
| DR   | Ffram; PF01506; HCV_Nsa; 1.  |                    |          |            |                         |
| DR   | Fram; PF00271; helicase_C; 1.  |                    |          |            |                         |
| DR   | Fram; PF00938; viral_RdRp; 1.  |                    |          |            |                         |
| KW   | SMART; SM00887; D8XDC; 1. Nonstructural protein; Polyprotein; RNA-directed RNA polymerase; Signal; Transferase.    |                    |          |            |                         |
| FT   | SIGNAL   |                    |          |            |                         |
| FT   | 55 83 20   |                    |          |            |                         |
| FT   | CHAIN 2 1 54   |                    |          |            |                         |
| FT   | CHAIN 84 300 NS2. NS3.   |                    |          |            |                         |
| FT   | CHAIN 301 931 NS3.   |                    |          |            |                         |
| FT   | CHAIN 932 985 NS3A.  |                    |          |            |                         |
| FT   | CHAIN 986 1245 NS3B.   |                    |          |            |                         |
| FT   | CHAIN 1247 1693 NS3A.  |                    |          |            |                         |
| FT   | CHAIN 1694 2284 NS3B.  |                    |          |            |                         |
| DR   | GO; GO_0016021; Cintegran to membrane; IEA.  |                    |          |            |                         |
| DR   | GO; GO_0019028; Cviral capsid; IEA.  |                    |          |            |                         |
| DR   | GO; GO_0019031; C:viral envelope; IEA.   |                    |          |            |                         |
| DR   | GO; GO_0005224; FATP binding; IEA.   |                    |          |            |                         |
| DR   | GO; GO_0008026; F:ATP dependent helicase activity; IEA.  |                    |          |            |                         |
| DR   | GO; GO_0003723; F:RNA binding; IEA.  |                    |          |            |                         |
| DR   | GO; GO_0003668; F:RNA-directed RNA polymerase activity; IEA.   |                    |          |            |                         |
| DR   | GO; GO_0008236; F:serine-type peptidase activity; IEA.   |                    |          |            |                         |
| DR   | GO; GO_0005198; F:structural molecule activity; IEA.   |                    |          |            |                         |
| DR   | GO; GO_0016740; F:transferase activity; IEA.   |                    |          |            |                         |
| DR   | GO; GO_000508; P:Proteolysis and peptidolysis; IEA.  |                    |          |            |                         |
| DR   | GO; GO_0006350; Pttranscription; IEA.  |                    |          |            |                         |
| DR   | GO; GO_0019079; P:viral genome replication; IEA.   |                    |          |            |                         |

DR GO; GO:0019087; P:viral transformation; IEA.  
 DR IPR00903; Cys Ser\_trypsin.  
 DR IPR01410; DEAD.  
 DR IPR002522; HCV\_capsid.  
 DR IPR002521; HCV\_core.  
 DR IPR002519; HCV\_env.  
 DR IPR002511; HCV\_NS1.  
 DR IPR002518; HCV\_NS2.  
 DR IPR00745; HCV\_NS4a.  
 DR IPR00745; HCV\_NS4b.  
 DR IPR002166; HCV\_NSSa.  
 DR IPR002165; HCV\_NSSb.  
 DR IPR001550; Helicase\_C.  
 DR IPR001551; Helicase\_C.  
 DR IPR004109; Peptidase\_C29.  
 DR IPR007095; RNA\_poli\_DS\_PS.  
 DR IPR007084; RNA\_poli\_PSVir.  
 DR PFAM; PRO1543; HCV\_core; 1.  
 DR PFAM; PRO1539; HCV\_env; 1.  
 DR PFAM; PRO1538; HCV\_NS2; 1.  
 DR PFAM; PRO1537; HCV\_NS3; 1.  
 DR PFAM; PRO1001; HCV\_NS4a; 1.  
 DR PFAM; PRO1001; HCV\_NS4b; 1.  
 DR PFAM; PRO1505; HCV\_NS5a; 1.  
 DR PFAM; PRO0271; Helicase\_C; 1.  
 DR PFAM; PRO0998; Viral\_RdRp; 1.  
 DR PRODOM; PD18062; HCV\_NS1; 1.  
 DR SMART; SM00487; DEXD\_C.  
 DR SEQUENCE; 3010 AA; 326880 MW; EED40E6A050B766 CRC64;

Query Match 96.2%; Score 1711; DB 12; length 3010;  
 Best Local Similarity 94.7%; Pred. No. 4..e-137;  
 Matches 323; Conservative 11; Mismatches 7; Indels 0; Gaps 0;

QY 1 VRGGDAAILLTCAVHPELIFDTIKLLAIFGPMVLQIAGITKPYFVRAOGLIRACMLV 60  
 866 VRRGRDAAILLTCAVHPELIFDTIKLLAIFGPMVLQIAGITKPYFVRAOGLIRACMLV 925  
 61 RKAAGHHYQVMAFMVLAUTGTYVYDHULPQLQWAHAGRDLAVAEVTSIDMEVKIT 120  
 926 WKAAGGHYQVMAFVKAALGTGTYVYDHULPQLRDLAHAAGRDLAVAEVPSVFSMETKIT 985

Db 121 WGADTAACGDIISGLPVSAARRGRELILGPAIDNPGQGMLLAPITASQTRGLGCIT 180  
 986 WGADTAACGDIISGLPVSAARRGRELILGPAIDNPGQGMLLAPITASQTRGLGCIT 1045  
 Db 181 SITGDRAKNOVGEVQVNSTATQSFELATCNGVWCTVFHAGSKTLLAGKGKPIQMYTWD 240  
 1046 SLTGDRAKNOVGEVQVNSTATQSFELATCNGVWCTVFHAGSKTLLAGKGKPIQMYTWD 1105  
 Qy 241 QDLVWQAPPARGASMPCTGGSDDYLVTRRHADIVPVRGDSRGSLISPPRVSYLKSS 300  
 Db 1106 QDLVWQAPPARGASMPCTGGSDDYLVTRRHADIVPVRGDSRGSLISPPRVSYLKSS 1165

Qy 301 CGPLICPSGHAVGIFRAVOTRGVAKAVDPFIPVSEMETWR 341  
 Db 1166 CGPLICPSGHAVGIFRAVOTRGVAKAVDPFIPVSEMETWR 1206

RESULT 15

P8996 PRELIMINARY; PRT; 3010 AA.

ID P89966  
 AC P89966;  
 DT 01-MAY-1997 (ITEMBREL. 03, Created)  
 DT 01-MAY-1997 (ITEMBREL. 03, Last sequence update)  
 DT 01-OCT-2003 (ITEMBREL. 25, Last annotation update)  
 DE RNA for polyprotein (Genome polyprotein).  
 OS Hepatitis C virus.  
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;  
 OC Recapiviruses.

OX NCBI\_TAXID=11103;  
 RN 11]  
 SEQUENCE FROM N.A.  
 RC STRAIN=type 1b;  
 RA Tanaka T.;  
 RL submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 STRAIN=type 1b;  
 RA TANAKA T.;  
 RT "TMORB\_n;"  
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases. COVERED BY A SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOFOLCAPSID COVERED BY A LIPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS: PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOFOLCAPSID IS A COMPLEX OF PROTEIN C AND MRENA (BY SIMILARITY).  
 CC EMBL; D89872; BASA4035.1; -.  
 DR PIR; A61196; A61196.  
 DR PIR; PQ0246; PQ0246.  
 DR PIR; PQ0246; PQ0246.  
 DR PIR; PQ0249; PQ0249.  
 DR HSSP; P26663; LXJP.  
 DR GO; GO:0016021; C:integral to membrane; IEA.  
 DR GO; GO:0019028; C:viral capsid; IEA.  
 DR GO; GO:0019031; C:viral envelope; IEA.  
 DR GO; GO:0005254; F:ATP binding; IEA.  
 DR GO; GO:0008026; F:ATP dependent helicase activity; IEA.  
 DR GO; GO:0003723; F:RNA binding; IEA.  
 DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.  
 DR GO; GO:0008236; F:serine-type peptidase activity; IEA.  
 DR GO; GO:000518; F:structural molecule activity; IEA.  
 DR GO; GO:0006740; F:transferase activity; IEA.  
 DR GO; GO:0006508; P:proteolysis and Peptidolysis; IEA.  
 DR GO; GO:0003650; P:transcription; IEA.  
 DR GO; GO:0019079; P:viral genome replication; IEA.  
 DR InterPro; IPR00903; Cys Ser\_trypsin.  
 DR InterPro; IPR01410; DEAD.  
 DR InterPro; IPR002521; HCV\_capsid.  
 DR InterPro; IPR02519; HCV\_env.  
 DR InterPro; IPR002531; HCV\_NS1.  
 DR InterPro; IPR002518; HCV\_NS2.  
 DR InterPro; IPR00745; HCV\_NS4a.  
 DR InterPro; IPR01490; HCV\_NS4b.  
 DR InterPro; IPR002868; HCV\_NS5a.  
 DR InterPro; IPR002166; HCV\_RdRp.  
 DR InterPro; IPR001650; Helicase\_C.  
 DR InterPro; IPR004109; Peptidase\_C29.  
 DR InterPro; IPR002511; HCV\_NS3.  
 DR InterPro; IPR007095; RNA\_poli\_DS\_PS.  
 DR InterPro; IPR007094; RNA\_poli\_PSVir.  
 DR Pfam; PF01543; HCV\_capsid; 1.  
 DR Pfam; PF01542; HCV\_core; 1.  
 DR Pfam; PF01539; HCV\_env; 1.  
 DR Pfam; PF01560; HCV\_NS1; 1.  
 DR Pfam; PF01538; HCV\_NS2; 1.  
 DR Pfam; PF02907; HCV\_NS3; 1.  
 DR Pfam; PF01006; HCV\_NS4a; 1.  
 DR Pfam; PF01001; HCV\_NS4b; 1.  
 DR Pfam; PF01506; HCV\_NS5a; 1.  
 DR Pfam; PF00271; Helicase\_C; 1.  
 DR Pfam; PF00398; Viral\_RdRp; 1.  
 DR PRODOM; PD18062; HCV\_NS1; 1.  
 DR SMART; SM00487; DEXD\_C.  
 KW Coat\_protein; Envelope\_protein; Glycoprotein; Nonstructural protein; Polyprotein; RNA-directed RNA polymerase; Transferase; Transmembrane.  
 SQ SEQUENCE 3010 AA; 327023 MW; B07BB9CFD811261 CRC64;

Query Match 96.2%; Score 1711; DB 12; Length 3010;  
 Best Local Similarity 95.3%; Pred. No. 4..e-137;  
 Matches 325; Conservative 9; Mismatches 7; Indels 0; Gaps 0;

QY 1 VRGGDAAILLTCAVHPELIFDTIKLLAIFGPMVLQIAGITKPYFVRAOGLIRACMLV 60

Db 866 ||||||| VGERDAILITCAVHPPELIFDIRGULALIGPLMVLQAGMTRVYFRAQGIRACMLV : ||||||| 925  
 QY 61 RKAAGGHYVOMAFMKLAALTGTYYDDHLPQLDVAHAGIRDLLAVAVEVTFISMEVKIT 1120  
 Db 926 RKVAGGHYVOMAFMKLAALTGTYYDHLPQLDVAHAGIRDLLAVAVEVTFISMEVKIT 985  
 QY 121 WGADTAACGDTISGLPVSAARRGRETLIGPDRNFGQGRLLAPITASQQTRELGICIT 180  
 Db 986 WGADTAACGDTISGLPVSAARRGRETLIGPDRNFGQGRLLAPITASQQTRELGICIT 1045  
 QY 181 SLTGRDKNOVGEVQVSTATOSFLATINGVCNTVPHAGSKTLAGPKPTOMYTND 240  
 Db 1046 SLTGRDKNOVGEVQVSTATOSFLATINGVCNTVPHAGSKTLAGPKPTOMYTND 1105  
 QY 241 QDLVWQAPGPGRSNTPCTCGSSSLVTRHADTVPRRGDRGSLSPPRPSYLGSS 300  
 Db 1106 QDLVWQAPGPGRSNTPCTCGSSSLVTRHADTVPRRGDRGSLSPPRPSYLGSS 1165  
 QY 301 GGPLCPSGHAVGIPRAAVCTRGYAKAVDPIPVESMETMR 341  
 Db 1166 GGPLCPSGHAVGIPRAAVCTRGYAKAVDPIPVESMETMR 1206

Search completed: May 6, 2004, 09:35:49  
 Job time : 32.8931 secs



Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:25:16 ; Search time 13:3837 Seconds  
 (without alignments)  
 1315.364 Million cell updates/sec

**Title:** US-10-650-585-14  
**Perfect score:** 1778  
**Sequence:** 1 VRGGRDAILLTCAVIRELPLI... RGVAKAVDTPVESNETMR 341

**Scoring table:** BLOSUM62  
**Gapop:** 10.0 , Gapext: 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
 Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries

**Database :**

- 1: Issued Patents AA:\*
- 2: /cgn2\_6/podata/2/1aa/5A\_COMB.pep:\*
- 3: /cgn2\_6/podata/2/1aa/6A\_COMB.pep:\*
- 4: /cgn2\_6/podata/2/1aa/6B\_COMB.pep:\*
- 5: /cgn2\_6/podata/2/1aa/pCTUS\_COMB.pep:\*
- 6: /cgn2\_6/podata/2/1aa/backfilest1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | Query Match | Length | DB ID               | Description        |
|------------|-------|-------------|--------|---------------------|--------------------|
| 1          | 1702  | 95.7        | 2201   | 4 US-09-539-601-6   | Sequence 6, Appli  |
| 2          | 1702  | 95.7        | 2201   | 4 US-09-539-601-15  | Sequence 15, Appli |
| 3          | 1702  | 95.7        | 3010   | 4 US-09-539-601-3   | Sequence 3, Appli  |
| 4          | 1702  | 95.7        | 3010   | 4 US-09-539-601-21  | Sequence 21, Appli |
| 5          | 1702  | 95.7        | 3010   | 4 US-09-539-601-27  | Sequence 27, Appli |
| 6          | 1698  | 95.5        | 1692   | 3 US-09-263-933-4   | Sequence 4, Appli  |
| 7          | 1698  | 95.5        | 1692   | 4 US-09-919-901-4   | Sequence 4, Appli  |
| 8          | 1698  | 95.5        | 2307   | 3 US-03-263-933-2   | Sequence 2, Appli  |
| 9          | 1698  | 95.5        | 2307   | 4 US-09-919-901-2   | Sequence 2, Appli  |
| 10         | 1695  | 95.3        | 1692   | 3 US-09-263-933-11  | Sequence 11, Appli |
| 11         | 1695  | 95.3        | 1692   | 4 US-09-919-901-11  | Sequence 11, Appli |
| 12         | 1695  | 95.3        | 2307   | 3 US-09-263-933-9   | Sequence 9, Appli  |
| 13         | 1695  | 95.3        | 2307   | 4 US-09-919-901-9   | Sequence 9, Appli  |
| 14         | 1695  | 95.3        | 3010   | 4 US-09-539-601-33  | Sequence 33, Appli |
| 15         | 1687  | 94.9        | 3010   | 3 US-09-019-416-3   | Sequence 3, Appli  |
| 16         | 1686  | 94.8        | 1692   | 3 US-09-263-933-18  | Sequence 18, Appli |
| 17         | 1686  | 94.8        | 1692   | 4 US-09-919-901-18  | Sequence 16, Appli |
| 18         | 1686  | 94.8        | 2307   | 3 US-09-263-933-16  | Sequence 16, Appli |
| 19         | 1685  | 94.8        | 2307   | 4 US-09-919-901-16  | Sequence 16, Appli |
| 20         | 1649  | 92.7        | 2013   | 1 US-08-324-977-12  | Sequence 12, Appli |
| 21         | 1649  | 92.7        | 2013   | 2 US-08-384-616-12  | Sequence 12, Appli |
| 22         | 1649  | 92.7        | 2013   | 3 US-08-904-686A-12 | Sequence 12, Appli |
| 23         | 1649  | 92.7        | 2013   | 3 US-09-315-855-12  | Sequence 12, Appli |
| 24         | 1649  | 92.7        | 2201   | 4 US-08-952-981A-2  | Sequence 2, Appli  |
| 25         | 1649  | 92.7        | 2620   | 2 US-08-324-977-32  | Sequence 32, Appli |
| 26         | 1649  | 92.7        | 2620   | 2 US-08-904-686A-32 | Sequence 32, Appli |
| 27         | 1649  | 92.7        | 2620   | 2 US-08-904-686A-32 | Sequence 32, Appli |

Sequence 32, Appli  
 Sequence 36, Appli  
 Sequence 36, Appli  
 Sequence 36, Appli  
 Sequence 2, Appli  
 Sequence 14, Appli  
 Sequence 2, Appli  
 Sequence 14, Appli  
 Sequence 2, Appli  
 Sequence 14, Appli  
 Sequence 14, Appli  
 Sequence 12, Appli  
 Sequence 12, Appli  
 Sequence 1, Appli  
 Sequence 1, Appli

RESULT 1  
 US-09-539-601-6  
 ; Sequence 6, Application US/953939601C  
 ; Patent No. 6630343  
 ; GENERAL INFORMATION  
 ; APPLICANT: Bartenschlager, Ralf FW  
 ; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System  
 ; FILE REFERENCE: all sequences  
 ; CURRENT APPLICATION NUMBER: US/09/539,601C  
 ; CURRENT FILING DATE: 2001-08-30  
 ; EARLIER APPLICATION NUMBER: 1999 15 178.4 GERMANY  
 ; EARLIER FILING DATE: 1999-04-03  
 ; NUMBER OF SEQ ID NOS: 51  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 5  
 ; LENGTH: 2201  
 ; TYPE: PRT  
 ; ORGANISM: Hepatitis C virus

US-09-539-601-6

Query Match Similarity 95.7%; Score 1702; DB 4; Length 2201;  
 Best Locl Similarity 94.1%; Pred. No. 4e-16; 9; Mismatches 11; Indels 0; Gaps 0;  
 Matches 321; Conservative 321;

QY |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
 1 VRGGRDAILLTCAVIRELPLI... RGVAKAVDTPVESNETMR 341  
 DB 57 VRGGRDAILLTCAVIRELPLI... RGVAKAVDTPVESNETMR 341  
 QY |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
 61 RKAGGIVQMAFMKLAALTGTIVDYLPIQDWAHAGRLAVAVEPVIFPSDEMKIT 120  
 DB 117 RKVAGGIVQMAFMKLAALTGTIVDYLPIQDWAHAGRLAVAVEPVIFPSDEMKIT 176  
 QY |||||:|||||:|||||:|||||:|||||:|||||:|||||:  
 121 WGADTAACGDTISGLPVSPARRGREGIELGPADNPEGGWRLLAPIAYSSQTORGILGCIT 180  
 DB 177 WGADTAACGDTISGLPVSPARRGREGIELGPADNPEGGWRLLAPIAYSSQTORGILGCIT 236  
 QY |||||:|||||:|||||:|||||:|||||:|||||:  
 181 SHTGRDKKQVGEVQVSTATOSFLATCVNGVCWTVPHGAGSKTLAGPKPITQMTND 240  
 DB 237 SLTGRDRNQVEGEVQVSTATOSFLATCVNGVCWTVPHGAGSKTLAGPKPITQMTND 296  
 QY |||||:|||||:|||||:|||||:|||||:  
 301 CGPLICSGHSAGVIFQAACVCRGVAKAVDTPVESNETMR 341  
 DB 357 GGPLCCPSGHAVGIFRAAVCTRGVAKAVDTPVESNETMR 397

RESULT 2

US-09-539-601-15

Sequence 15, Application US/09539601C  
; Patent No. 6630343  
; GENERAL INFORMATION:  
; APPLICANT: Bartenschlager, Ralf FW  
; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System  
; FILE REFERENCE: all sequences  
; CURRENT APPLICATION NUMBER: US/09/539,601C  
; CURRENT FILING DATE: 2001-08-30  
; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY  
; EARLIER FILING DATE: 1999-04-03  
; NUMBER OF SEQ ID NOS: 51  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO: 15  
; LENGTH: 2201  
; TYPE: PRT  
; ORGANISM: Hepatitis C virus  
; US-09-539-601-15

Query Match 95.7%; Score 1702; DB 4; Length 2201;  
Best Local Similarity 94.1%; Pred. No. 6\_3e-164; Mismatches 11; Indels 0; Gaps 0;  
Matches 321; Conservative

Qy 1 VRGGRDAVILITCAHPBLIFITKLLIAFGLPLMVLQAGITKPYFRAQGIRACMLV 60  
Db 57 VRGGRDAVILITCAHPBLIFITKLLIAFGLPLMVLQAGITKPYFRAQGIRACMLV 116  
Qy 61 RKAGAGHGVQAMFMKLAALTGTIVYDHITPLDWAHAGIRDIAVAVPVPFSDMEVKIT 120  
Db 117 RKVAGGHVQAMLMKLAALTGTIVYDHITPLDWAHAGIRDIAVAVPVPFSDMEVKIT 176  
Qy 121 WGADTAAGDIISGLPVSARRREILGPADNFGQGRLLAPITASQQTREGLGICIT 180  
Db 177 WGADTAAGDIISGLPVSARRREIHGPADSLLEGQGRLLAPITASQQTREGLGICIT 236  
Qy 181 SLTGRDKRQVEQEVQVSTATQSFLATCUNGCVWTVFHAGSKTLASPKGPIQOMYNTD 240  
Db 237 SLTGRDRNQVEEVQVSTATQSFLATCUNGCVWTVFHAGSKTLASPKGPIQOMYNTD 296  
Qy 241 QDLVGWQAPPGARSMTCTCGSSDLYLVRHADVIPTVRRGDSRSGLSISPRPSYLGSS 300  
Db 297 QDLVGWQAPPGARSMTCTCGSSDLYLVRHADVIPTVRRGDSRSGLSISPRPSYLGSS 356

Qy 301 GGPLCPSPGHAVGIFRAAVCTRGVAKAVDFVPPVESMETMR 341  
Db 357 GGPLCPSPGHAVGIFRAAVCTRGVAKAVDFVPPVESMETMR 397

RESULT 3

US-09-539-601-3

Sequence 3, Application US/09539601C  
; Patent No. 6630343  
; GENERAL INFORMATION:  
; APPLICANT: Bartenschlager, Ralf FW  
; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System  
; FILE REFERENCE: all sequences  
; CURRENT APPLICATION NUMBER: US/09/539,601C  
; CURRENT FILING DATE: 2001-08-30  
; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY  
; EARLIER FILING DATE: 1999-04-03  
; NUMBER OF SEQ ID NOS: 51  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO: 21  
; LENGTH: 3010  
; TYPE: PRT  
; ORGANISM: Hepatitis C virus  
; US-09-539-601-21

Query Match 95.7%; Score 1702; DB 4; Length 3010;  
Best Local Similarity 94.1%; Pred. No. 6\_3e-164; Mismatches 11; Indels 0; Gaps 0;  
Matches 321; Conservative

Qy 1 VRGGRDAVILITCAHPBLIFITKLLIAFGLPLMVLQAGITKPYFRAQGIRACMLV 60  
Db 866 VRGGRDAVILITCAHPBLIFITKLLIAFGLPLMVLQAGITKPYFRAQGIRACMLV 925  
Qy 61 RKAGAGHGVQAMFMKLAALTGTIVYDHITPLDWAHAGIRDIAVAVPVPFSDMEVKIT 120  
Db 926 RKVAGGHVQAMLMKLAALTGTIVYDHITPLDWAHAGIRDIAVAVPVPFSDMEVKIT 985  
Qy 121 WGADTAAGDIISGLPVSARRREILGPADNFGQGRLLAPITASQQTREGLGICIT 180  
Db 986 WGADTAAGDIISGLPVSARRREIHGPADSLLEGQGRLLAPITASQQTREGLGICIT 1045  
Qy 181 SLTGRDKRQVEQEVQVSTATQSFLATCUNGCVWTVFHAGSKTLASPKGPIQOMYNTD 240  
Db 1046 SLTGRDRNQVEEVQVSTATQSFLATCUNGCVWTVFHAGSKTLASPKGPIQOMYNTD 1105  
Qy 241 QDLVGWQAPPGARSMTCTCGSSDLYLVRHADVIPTVRRGDSRSGLSISPRPSYLGSS 300  
Db 1106 QDLVGWQAPPGARSMTCTCGSSDLYLVRHADVIPTVRRGDSRSGLSISPRPSYLGSS 1165

Query Match 95.7%; Score 1702; DB 4; Length 3010;  
Best Local Similarity 94.1%; Pred. No. 6\_3e-164; Mismatches 11; Indels 0; Gaps 0;  
Matches 321; Conservative

Qy 301 GGPLCPSPGHAVGIFRAAVCTRGVAKAVDFVPPVESMETMR 341  
Db 1166 GGPLCPSPGHAVGIFRAAVCTRGVAKAVDFVPPVESMETMR 1206

RESULT 4

US-09-539-601-21

Sequence 21, Application US/09539601C  
; Patent No. 6630343  
; GENERAL INFORMATION:  
; APPLICANT: Bartenschlager, Ralf FW  
; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System  
; FILE REFERENCE: all sequences  
; CURRENT APPLICATION NUMBER: US/09/539,601C  
; CURRENT FILING DATE: 2001-08-30  
; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY  
; EARLIER FILING DATE: 1999-04-03  
; NUMBER OF SEQ ID NOS: 51  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO: 21  
; LENGTH: 3010  
; TYPE: PRT  
; ORGANISM: Hepatitis C virus  
; US-09-539-601-21

Query Match 95.7%; Score 1702; DB 4; Length 3010;  
Best Local Similarity 94.1%; Pred. No. 6\_3e-164; Mismatches 11; Indels 0; Gaps 0;  
Matches 321; Conservative

Qy 1 VRGGRDAVILITCAHPBLIFITKLLIAFGLPLMVLQAGITKPYFRAQGIRACMLV 60  
Db 866 VRGGRDAVILITCAHPBLIFITKLLIAFGLPLMVLQAGITKPYFRAQGIRACMLV 925  
Qy 61 RKAGAGHGVQAMFMKLAALTGTIVYDHITPLDWAHAGIRDIAVAVPVPFSDMEVKIT 120  
Db 926 RKVAGGHVQAMLMKLAALTGTIVYDHITPLDWAHAGIRDIAVAVPVPFSDMEVKIT 985  
Qy 121 WGADTAAGDIISGLPVSARRREILGPADNFGQGRLLAPITASQQTREGLGICIT 180  
Db 986 WGADTAAGDIISGLPVSARRREIHGPADSLLEGQGRLLAPITASQQTREGLGICIT 1045  
Qy 181 SLTGRDKRQVEQEVQVSTATQSFLATCUNGCVWTVFHAGSKTLASPKGPIQOMYNTD 240  
Db 1046 SLTGRDRNQVEEVQVSTATQSFLATCUNGCVWTVFHAGSKTLASPKGPIQOMYNTD 1105  
Qy 241 QDLVGWQAPPGARSMTCTCGSSDLYLVRHADVIPTVRRGDSRSGLSISPRPSYLGSS 300  
Db 1106 QDLVGWQAPPGARSMTCTCGSSDLYLVRHADVIPTVRRGDSRSGLSISPRPSYLGSS 1165

RESULT 5  
US-09-539-601-27  
; Sequence 27, Application US/09539601C  
; Pattern No. 663043  
; GENERAL INFORMATION:  
; APPLICANT: Barrenschlager, Ralf FW  
; TITLE OF INVENTION: Hepatitis C Virus Cell Culture System  
; FILE REFERENCE: all sequences  
; CURRENT APPLICATION NUMBER: US/09/539,601C  
; CURRENT FILING DATE: 2001-08-30  
; EARLIER APPLICATION NUMBER: 199 15 178.4 GERMANY  
; EARLIER FILING DATE: 1999-04-03  
; NUMBER OF SEQ ID NOS: 51  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 27  
; LENGTH: 3010  
; TYPE: PRT  
; ORGANISM: Hepatitis C virus  
US-09-539-601-27

Query Match 95.7%; Score 1702; DB 4; Length 3010;  
Best Local Similarity 94.1%; Pred. No. 6.3e-164;  
Matches 321; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

Qy 1 VRGGRDAILITTCAVPHELIPDITKLLIAIFGPMLVQAGITKVYPFVRAGLIRACMLV 60  
Db 866 VRGGRDAVILITCAIHPELIFTKILLALIGPLMVLAGITKVYPFVRAGLIRACMLV 925

Qy 61 RKAAGGHYQMAFMKLAALTGTYYDHPLQDWAHAGLIRDIAVAVSPVFSDEVKIT 120  
Db 926 RKGAGHGYQMAFMKLAALTGTYYDHPLQDWAHAGLIRDIAVAVSPVFSDEVKIT 985

Qy 121 WGADTAACCDITISGLPVSARRGREBILGFLGPMVLAQGWRLLAPITASQQTGELCIT 180  
Db 986 WGADTAACCDITISGLPVSARRGREBILGFLGPMVLAQGWRLLAPITASQQTGELCIT 1045

Qy 181 SLGRDRDKNQVEGEVQVSTATOSFLATCNGVWTFVFGAGSKTLACPKGPRTQMYINVD 240  
Db 1046 SLGRDRDKNQVEGEVQVSTATOSFLATCNGVWTFVFGAGSKTLACPKGPRTQMYINVD 1105

Qy 241 QDVGWQAPPGRASMTPCCGSSDLYLVTRADYVTPRRGSRGSLSPRVSYINGSS 300  
Db 1106 QDVGWQAPPGRASLTPTCGSSDLYLVTRADYVTPRRGSRGSLSPRVSYINGSS 1165

Qy 301 GGPLCPSPSHAVGTFRAAVCTRGVAKAVDFPIPVESMETMR 341  
Db 1166 GGPLCPSPSHAVGTFRAAVCTRGVAKAVDFPVVESMETMR 1206

RESULT 6  
US-09-263-933-4  
; Sequence 4, Application US/09263933  
; Pattern No. 628040  
; GENERAL INFORMATION:  
; APPLICANT: Potts, Karen E.  
; APPLICANT: Jackson, Roberta L.  
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
; FILE REFERENCE: 0125-005A  
; CURRENT APPLICATION NUMBER: US/09/919, 901  
; CURRENT FILING DATE: 2001-08-02  
; PRIOR APPLICATION NUMBER: 09/263, 933  
; PRIOR FILING DATE: 1999-02-08  
; PRIORITY NUMBER: 09/129, 611  
; PRIOR FILING DATE: 1998-08-05  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 1692  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE: :  
; OTHER INFORMATION: :

US-09-919-901-4  
Query Match 95.5%; Score 1698; DB 4; Length 1692;  
Best Local Similarity 94.1%; Pred. No. 6.9e-164;  
Matches 320; Conservative 12; Mismatches 8; Indels 0; Gaps 0;

Qy 2 RGRDAILITTCAVPHELIPDITKLLIAIFGPMLVQAGITKVYPFVRAGLIRACMLV 61  
Db 146 RGRDAILITTCAVPHELIPDITKLLIAIFGPMLVQAGITKVYPFVRAGLIRACMLV 205

Qy 62 KAAGGHYQMAFMKLAALTGTYYDHPLQDWAHAGLIRDIAVAVSPVFSDEVKITW 121  
Db 206 KGAGHGYQMAFMKLAALTGTYYDHPLQDWAHAGLIRDIAVAVSPVFSDEVKITW 265

Qy 122 GADTAACCDITISGLPVSARRGREBILGFLGPMVLAQGWRLLAPITASQQTGELCITS 181  
Db 266 GADTAACCDITISGLPVSARRGREBILGFLGPMVLAQGWRLLAPITASQQTGELCITS 325

Qy 182 LTGRDRDKNQVEGEVQVSTATOSFLATCNGVWTFVFGAGSKTLACPKGPRTQMYINVD 241  
Db 326 LTGRDRDKNQVEGEVQVSTATOSFLATCNGVWTFVFGAGSKTLACPKGPRTQMYINVD 385

Qy 302 GPLICRGSGHAGVIFRAAVCTRGVAKAVDFPIPVESMETMR 341  
Db 446 GPLICRGSGHAGVIFRAAVCTRGVAKAVDFPVVESMETMR 485

RESULT 7  
US-09-919-901-4  
; Sequence 4, Application US/09919901  
; Pattern No. 6599738  
; GENERAL INFORMATION:  
; APPLICANT: Potts, Karen E.  
; APPLICANT: Jackson, Roberta L.  
; APPLICANT: Patrick, Amy K.  
; TITLE OF INVENTION: OP INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE  
; FILE REFERENCE: 0125-005A  
; CURRENT APPLICATION NUMBER: US/09/919, 901  
; CURRENT FILING DATE: 2001-08-02  
; PRIOR APPLICATION NUMBER: 09/263, 933  
; PRIOR FILING DATE: 1999-02-08  
; PRIORITY NUMBER: 09/129, 611  
; PRIOR FILING DATE: 1998-08-05  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 1692  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE: :  
; OTHER INFORMATION: :

US-09-919-901-4  
Query Match 95.5%; Score 1698; DB 4; Length 1692;  
Best Local Similarity 94.1%; Pred. No. 6.9e-164;  
Matches 320; Conservative 12; Mismatches 8; Indels 0; Gaps 0;

Qy 2 RGRDAILITTCAVPHELIPDITKLLIAIFGPMLVQAGITKVYPFVRAGLIRACMLV 61  
Db 146 RGRDAILITTCAVPHELIPDITKLLIAIFGPMLVQAGITKVYPFVRAGLIRACMLV 205

Qy 62 KAAGGHYQMAFMKLAALTGTYYDHPLQDWAHAGLIRDIAVAVSPVFSDEVKITW 121  
Db 206 KGAGHGYQMAFMKLAALTGTYYDHPLQDWAHAGLIRDIAVAVSPVFSDEVKITW 265

Qy 122 GADTAACCDITISGLPVSARRGREBILGFLGPMVLAQGWRLLAPITASQQTGELCITS 181  
Db 266 GADTAACCDITISGLPVSARRGREBILGFLGPMVLAQGWRLLAPITASQQTGELCITS 325

Qy 182 LTGRDRDKNQVEGEVQVSTATOSFLATCNGVWTFVFGAGSKTLACPKGPRTQMYINVD 241

RESULT 8  
 US-09-263-933-2  
 ; Sequence 2, Application US/09263933  
 ; Patent No. 6280940  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Potts, Karen E.  
 ; APPLICANT: Jackson, Roberta L.  
 ; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
 ; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE  
 ; FILE REFERENCE: 0125-0005A  
 ; CURRENT APPLICATION NUMBER: US/09/919, 901  
 ; CURRENT FILING DATE: 2001-08-02  
 ; PRIOR APPLICATION NUMBER: 09/263, 933  
 ; PRIOR FILING DATE: 1999-02-08  
 ; PRIOR APPLICATION NUMBER: 09/129, 611  
 ; NUMBER OF SEQ ID NOS: 33  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO: 2  
 ; LENGTH: 2307  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; OTHER INFORMATION:  
 ; FEATURE:  
 ; LENGTH: 2307  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; US-09-263-933-2  
 Query Match 95.3%; Score 1698; DB 3; Length 2307;  
 Best Local Similarity 94.1%; Pred. No. 1..1e-163;  
 Matches 320; Conservative 12; Mismatches 8; Indels 0; Gaps 0;  
 Query 2 RGGRDAILLLTCVPHPELFDITKILLAIIFGPMLVQAGITKPYFVRQGLRACMLVR 61  
 Db 238 RGGRDAILLLTCVPHPELFDITKILLAIIFGPMLVQAGITKPYFVRQGLRACMLVR 297  
 Query 62 KAAGGHIVQVAFMKLAALGTYVVDHPLQDWAHAGIRDLAYAVEPVIFSNEVKITW 121  
 Db 298 KAAGGHIVQVAFMKLAALGTYVVDHPLQDWAHAGIRDLAYAVEPVIFSNEVKITW 357  
 Query 122 GADTAACGDIISGIRVSARRGREILINGPDLNFGCGWRILLAPITASQTRGILGCITS 181  
 Db 418 LTGRDKRNQVEGVQVSTATOSFLATCNGVCWVYHAGSKTLAGPKGPITQMYTNQD 477  
 Query 242 DLVGMQAPPGRASMPCTCGSSDLYLVRHADVIPVRRGDSRGSLSLSPRPSYLGSS 301  
 Db 478 DLVGMQAPPGRASLPCTCGSSDLYLVRHADVIPVRRGDSRGSLSLSPRPSYLGSS 537  
 Query 182 LTGRDKRNQGEVQVSTATOSFLATCNGVCWVYHAGSKTLAGPKGPITQMYTNQD 241  
 Db 538 GPLICPSGHAVGIFRAAVCTRGVAKAVDPVPSMETMR 577  
 RESULT 9  
 US-09-919-901-2  
 ; Sequence 2, Application US/09919901  
 ; Patent No. 6599738  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Potts, Karen E.  
 ; APPLICANT: Jackson, Roberta L.  
 ; APPLICANT: Patick, Amy K.  
 ; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
 ; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE  
 Query Match 95.3%; Score 1695; DB 3; Length 1692;  
 Best Local Similarity 93.8%; Pred. No. 1..4e-163; Indels 0; Gaps 0;  
 Matches 319; Conservative 13; Mismatches 8; Indels 0; Gaps 0;  
 Query Match 95.3%; Score 1695; DB 3; Length 1692;  
 Best Local Similarity 93.8%; Pred. No. 1..4e-163; Indels 0; Gaps 0;  
 Matches 319; Conservative 13; Mismatches 8; Indels 0; Gaps 0;

QY 2 RGGRDAILLTCAVHPLEIFDITKLLAIFGPMVLQAGITKVYFVRAGQLIRACMLVR 61  
Db 146 RGGRDAILLTCAVHPLEIFDITKLLAIFGPMVLQAGITKVYFVRAGQLIRACMLVR 205

QY 62 KAAGGHYVQMAFMKLAALTGTYYVDLHPLQDWAHAGRLAVAPVTFQVRAQGLIACMLVR 121  
Db 205 KVAGGHYVQMAFMKLAALTGTYYVDLHPLQDWAHAGRLAVAPVTFQVRAQGLIACMLVR 265

QY 122 GADTAACGDIISGLPVSARRGKETILGPDSLGWRLLAPIAYSQTRGLGCITS 181  
Db 266 GADTAACGDIISGLPVSARRGKETILGPDSLGWRLLAPIAYSQTRGLGCITS 325

QY 182 LTGRDKNOVEGEVQVSTATOSFLATCNGVWTVFHAGSKTLAGPKGPIQMTNVDQ 241  
Db 326 LTGRDKNOVEGEVQVSTATOSFLATCNGVWTVFHAGSKTLAGPKGPIQMTNVDQ 385

QY 242 DLVGWQAPPGARSMTPTCGSSDLVLTTRHADVIPVRGGDSRGSLSLSPRVSYLKSG 301  
Db 386 DLVGWQAPPGARSMTPTCGSSDLVLTTRHADVIPVRGGDSRGSLSLSPRVSYLKSG 445

QY 302 GPLLCPSGHAVGIFRAAVCTRGAKAVDFPVVESMETMR 341  
Db 446 GPLLCPSGHAVGIFRAAVCTRGAKAVDFPVVESMETMR 485

RESULT 11  
US-09-919-901-11  
; Sequence 11, Application US/09919901

Patent No. 6599738  
; GENERAL INFORMATION:  
; APPLICANT: Potts, Karen E.  
; APPLICANT: Jackson, Roberta L.  
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE  
; FILE REFERENCE: 0125-005A  
; CURRENT APPLICATION NUMBER: US/09/919,901  
; CURRENT FILING DATE: 2001-08-02  
; PRIOR APPLICATION NUMBER: 09/1263,933  
; PRIOR FILING DATE: 1999-02-08  
; PRIOR FILING DATE: 09/129,611  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO: 9  
; LENGTH: 2307  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; US-09-263-933-9

Query Match 95.3%; Score 1695; DB 3; Length 2307;  
Best Local Similarity 93.8%; Pred. No. 2.2e-163; Mismatches 8; Indels 0; Gaps 0;  
Matches 319; Conservative 13; MisMatches 8; Indels 0; Gaps 0;

QY 2 RGGRDAILLTCAVHPLEIFDITKLLAIFGPMVLQAGITKVYFVRAGQLIRACMLVR 61  
Db 238 RGGRDAILLTCAVHPLEIFDITKLLAIFGPMVLQAGITKVYFVRAGQLIRACMLVR 297

QY 62 KAAGGHYVQMAFMKLAALTGTYYVDLHPLQDWAHAGRLAVAPVTFQVRAQGLIACMLVR 121  
Db 298 KVAGGHYVQMAFMKLAALTGTYYVDLHPLQDWAHAGRLAVAPVTFQVRAQGLIACMLVR 357

QY 122 GADTAACGDIISGLPVSARRGKETILGPDSLGWRLLAPIAYSQTRGLGCITS 181  
Db 358 GADTAACGDIISGLPVSARRGKETILGPDSLGWRLLAPIAYSQTRGLGCITS 417

QY 182 DLVGWQAPPGARSMTPTCGSSDLVLTTRHADVIPVRGGDSRGSLSLSPRVSYLKSG 301  
Db 418 LTGRDKNOVEGEVQVSTATOSFLATCNGVWTVFHAGSKTLAGPKGPIQMTNVDQ 477

QY 242 DLVGWQAPPGARSMTPTCGSSDLVLTTRHADVIPVRGGDSRGSLSLSPRVSYLKSG 301  
Db 478 DLVGWQAPPGARSMTPTCGSSDLVLTTRHADVIPVRGGDSRGSLSLSPRVSYLKSG 537

RESULT 13  
US-09-919-901-9  
; Sequence 9, Application US/09919901  
; Patent No. 6599738  
; GENERAL INFORMATION:  
; APPLICANT: Potts, Karen E.  
; APPLICANT: Jackson, Roberta L.  
; TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT  
; TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE  
; FILE REFERENCE: 0125-005A  
; CURRENT APPLICATION NUMBER: US/09/919,901  
; CURRENT FILING DATE: 2001-08-02

PRIOR APPLICATION NUMBER: 09/263, 933  
PRIOR FILING DATE: 1999-02-08  
PRIOR APPLICATION NUMBER: 09/129, 611  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO: 9  
LENGTH: 2307  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION:  
US-09-919-901-9

Query Match Score 95.3%; DB 4; Length 2307;  
Best Local Similarity 93.8%; Pred. No. 2.2e-163; Matches 319; Conservative 13; Mismatches 8; Indels 0; Gaps 0;

QY 2 RGGDAAILTCATPHELIFDITKLLAIFGPMVNLQAGITKPYFRAOGLIRACMLV 61  
Db 238 RGGRDAAILMCATPHELIFDITKLLAIFGPMVNLQAGITKPYFRAOGLIRACMLV 297  
Qy 62 KAAAGHYYVQMAFMKLAALTGTYYDHPLQDMAHAGIRDALAVEPVIFPSMEVKITW 121  
Db 298 KVAGGHHYVQMAFMKLGALTGTYYNHPLRWAHAGIRDALAVEPVIFPSMETKIVT 357  
Qy 122 GADTAACGDIISGLPVSAARRGRETLIGPADNFGQWRLLAPITAYSOQTRLGICITS 181  
Db 358 GADTAACGDIISGLPVSAARRGRETLIGPADNFGQWRLLAPITAYSOQTRLGICITS 417  
Qy 182 LTGDKDNOEGEVOVSTATOSLATCNGVCMVFHAGSKTLAGRKGPIOMYTVQD 241  
Db 418 LTGRKKNQVGEVQVSTATOSLATCNGVCMVFHAGSKTLAGRKGPIOMYTVQD 477  
Qy 242 DLVHQAPPGARSMTPTCGSSDLYLVRHADIVPVERRGDSRGSLSLSPRVSYKSS 301  
Db 478 DLVHQAPPGARSMTPTCGSSDLYLVRHADIVPVERRGDSRGSLSLSPRVSYKSS 537

Qy 302 GPLIOPSGHVGIFRAAVCTRGVAKAVDPFPVPSMETMR' 341  
Db 538 GPLIOPSGHVGIFRAAVCTRGVAKAVDPFPVPSMETMR' 577

---

RESULT 14

US-09-539-601-33 Application US/09539601C  
Sequence 33; Application US/09539601C  
Patent No. 663043  
GENERAL INFORMATION:  
APPLICANT: Barrensclager, Ralf FW  
TITLE OF INVENTION: Hepatitis C Virus Cell Culture System  
FILE REFERENCE: all sequences  
CURRENT APPLICATION NUMBER: US/09/539, 601C  
CURRENT FILING DATE: 2001-08-30  
EARLIER APPLICATION NUMBER: 199 15 178-4 GERMANY  
EARLIER FILING DATE: 1999-04-03  
NUMBER OF SEQ ID NOS: 51  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO: 33  
LENGTH: 3010  
TYPE: PRT  
ORGANISM: Hepatitis C virus  
US-09-014-416-3

Query Match Score 94.9%; DB 3; Length 3010;  
Best Local Similarity 93.3%; Pred. No. 2.2e-162; Matches 318; Conservative 12; Mismatches 11; Indels 0; Gaps 0;

Qy 1 VRGGDAAILTCATPHELIFDITKLLAIFGPMVNLQAGITKPYFRAOGLIRACMLV 60  
Db 866 VRGGDAAILTCATPHELIFDITKLLAIFGPMVNLQAGITKPYFRAOGLIRACMLV 925  
Qy 61 RKAAGHYYVQMAFMKLAALTGTYYDHPLQDMAHAGIRDALAVEPVIFPSMEVKIT 120  
Db 926 RKVAGGHHYVQMAFMKLGALTGTYYNHPLRWAHAGIRDALAVEPVIFPSMETKIVT 985  
Qy 121 WGADTAACGDIISGLPVSAARRGRETLIGPADNFGQWRLLAPITASQQTRLGICITS 180  
Db 986 WGADTAACGDIISGLPVSAARRGRETLIGPADNFGQWRLLAPITASQQTRLGICITS 1045  
Qy 181 SLTGDKNOEGEVOVSTATOSLATCNGVCMVFHAGSKTLAGRKGPIOMYTVQD 240  
Db 1046 SLTGDKNOEGEVOVSTATOSLATCNGVCMVFHAGSKTLAGRKGPIOMYTVQD 1105  
Qy 241 ODLVHQAPPGARSMTPTCGSSDLYLVRHADIVPVERRGDSRGSLSLSPRVSYKSS 300  
Db 1106 LIDLVHQAPPGARSMTPTCGSSDLYLVRHADIVPVERRGDSRGSLSLSPRVSYKSS 1165  
Qy 301 GGPLIOPSGHVGIFRAAVCTRGVAKAVDPFPVPSMETMR' 341  
Db 1166 GGPLIOPSGHVGIFRAAVCTRGVAKAVDPFPVPSMETMR' 1206

Search completed: May 6, 2004, 09:39:05

Fri May 7 13:37:05 2004

Job time : 14.3837 secs

us-10-650-585-14.rai

Page 7





Db 181 SLTGRDKNOVEGEVQVSTATQSFLATCNGVCWTVFHAGSKTLAGPKGPIQMTNVD 240 ; PRIOR FILING DATE: 2000-12-15  
 Qy 241 QDLVGWQAPPGRASMPCTCGSSDLYLVRHADYIPVRERGDRGSLSPRPSYLKSS 300 ; NUMBER OF SEQ ID NOS: 21  
 ; SOFTWARE: FastSEQ for Windows Version 4.0  
 Db 241 QDLVGWQAPPGRASMPCTCGSSDLYLVRHADYIPVRERGDRGSLSPRPSYLKSS 300 ; SEQ ID NO 13  
 ; LENGTH: 352  
 ; TYPE: PRT  
 Qy 301 GGPLICPSGHAVGIFRAVCTRGVAKAVDPIPESMETMR 341 ; ORGANISM: HCV  
 Db 301 GGPLICPSGHAVGIFRAVCTRGVAKAVDPIPESMETMR 341 ;  
 RESULT 2 ;  
 US-10-650-585-14 ;  
 ; Sequence 14. Application US/10650585  
 ; Publication No. US20040077066A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Boehringer Ingelheim (Canada) Ltd. Protease  
 ; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease  
 ; FILE REFERENCE: 13/082  
 ; CURRENT APPLICATION NUMBER: US10/650, 585  
 ; CURRENT FILING DATE: 2003-08-28  
 ; PRIOR APPLICATION NUMBER: US/10/017, 736A  
 ; PRIOR FILING DATE: 2001-12-14  
 ; PRIOR APPLICATION NUMBER: 60/256, 031  
 ; PRIOR FILING DATE: 2000-12-15  
 ; NUMBER OF SEQ ID NOS: 21  
 ; SOFTWARE: FastSEQ for Windows Version 4.0  
 ; SEQ ID NO 14  
 ; LENGTH: 341  
 ; TYPE: PRT  
 ; ORGANISM: HCV  
 ; US-10-650-585-14  
 Query Match 100.0%; Score 1778; DB 16; Length 341;  
 Best local Similarity 100.0%; Pred. No. 2.9e-174; Mismatches 0; Indels 0; Gaps 0;  
 Matches 341; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 VRGGDAAILLTCAVHPPELFIDITKLLAIFGPMVLQAGITKPYFVRAQSLIRACMV 60 ;  
 Db 12 VRGGDAAILLTCAVHPPELFIDITKLLAIFGPMVLQAGITKPYFVRAQSLIRACMV 60 ;  
 Qy 61 RKAAGHYYVOMAFMKLAALGTGYVYDHLPQLQWAHAGIRDLAVERVIFSDMEVKIT 120 ;  
 Db 72 RKAAGHYYVOMAFMKLAALGTGYVYDHLPQLQWAHAGIRDLAVERVIFSDMEVKIT 131 ;  
 Qy 121 WGADTAACGDIISGLPVSARRGREILLGPADNFGQWRLLAPITAYSQQTRLGCLIT 180 ;  
 Db 132 WGADTAACGDIISGLPVSARRGREILLGPADNFGQWRLLAPITAYSQQTRLGCLIT 191 ;  
 Qy 181 SLTGRDKNOVEGEVQVSTATQSFLATCNGVCWTVFHAGSKTLAGPKGPIQMTNVD 240 ;  
 Db 192 SLTGRDKNOVEGEVQVSTATQSFLATCNGVCWTVFHAGSKTLAGPKGPIQMTNVD 251 ;  
 Qy 241 QDLVGWQAPPGRASMPCTCGSSDLYLVRHADYIPVRERGDRGSLSPRPSYLKSS 300 ;  
 Db 252 QDLVGWQAPPGRASMPCTCGSSDLYLVRHADYIPVRERGDRGSLSPRPSYLKSS 311 ;  
 Qy 301 GGPLICPSGHAVGIFRAVCTRGVAKAVDPIPESMETMR 341 ;  
 Db 312 GGPLICPSGHAVGIFRAVCTRGVAKAVDPIPESMETMR 352 ;  
 RESULT 4 ;  
 US-10-650-585-13 ;  
 ; Sequence 13, Application US/10650585  
 ; Publication No. US20040077066A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Boehringer Ingelheim (Canada) Ltd. Protease  
 ; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease  
 ; FILE REFERENCE: 13/082  
 ; CURRENT APPLICATION NUMBER: US/10/650, 585  
 ; PRIOR APPLICATION NUMBER: US/10/017, 736A  
 ; PRIOR FILING DATE: 2001-12-14  
 ; PRIOR APPLICATION NUMBER: 60/256, 031  
 ; PRIOR FILING DATE: 2000-12-15  
 ; NUMBER OF SEQ ID NOS: 21  
 ; SOFTWARE: FastSEQ for Windows Version 4.0  
 ; SEQ ID NO 13  
 ; LENGTH: 352  
 ; TYPE: PRT  
 ; ORGANISM: HCV  
 ; US-10-650-585-13  
 Query Match 100.0%; Score 1778; DB 16; Length 352;  
 Best local Similarity 100.0%; Pred. No. 2.9e-174; Mismatches 0; Indels 0; Gaps 0;  
 Matches 341; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 VRGGDAAILLTCAVHPPELFIDITKLLAIFGPMVLQAGITKPYFVRAQSLIRACMV 60 ;  
 Db 12 VRGGDAAILLTCAVHPPELFIDITKLLAIFGPMVLQAGITKPYFVRAQSLIRACMV 71 ;  
 Qy 61 RKAAGHYYVOMAFMKLAALGTGYVYDHLPQLQWAHAGIRDLAVERVIFSDMEVKIT 120 ;  
 Db 72 RKAAGHYYVOMAFMKLAALGTGYVYDHLPQLQWAHAGIRDLAVERVIFSDMEVKIT 131 ;  
 Qy 121 WGADTAACGDIISGLPVSARRGREILLGPADNFGQWRLLAPITAYSQQTRLGCLIT 180 ;

RESULT 5  
 US-10-017-736-12  
 ; Sequence 12, Application US/10017736  
 ; Publication No. US20020192640A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Behringwer Ingelheim (Canada) Ltd.  
 ; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease  
 ; FILE REFERENCE: 13/082  
 ; CURRENT APPLICATION NUMBER: US/10/017,736  
 ; CURRENT FILING DATE: 2001-12-14  
 ; PRIOR APPLICATION NUMBER: 60/256,031  
 ; PRIOR FILING DATE: 2000-12-15  
 ; NUMBER OF SEQ ID NOS: 21  
 ; SOFTWARE: FastSEQ for Windows Version 4.0  
 ; SEQ ID NO 12  
 ; LENGTH: 380  
 ; TYPE: PRT  
 ; ORGANISM: HCV  
 ; US-10-017-736-12

Query Match 100.0%; Score 1778; DB 13; Length 380;  
 Best Local Similarity 100.0%; Pred. No. 3.2e-174; Matches 341; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Gaps 0;

QY 1 VRGGDAAILLTCAVHPELIFDITKLLIAIFGPLMVLQAGITKPYFVRAQGLIRACMV 60  
 Db 40 VRGGDAAILLTCAVHPELIFDITKLLIAIFGPLMVLQAGITKPYFVRAQGLIRACMV 99  
 QY 61 RKAAGGHHYVOMAFMKLAALGTYVYDHLPLODWAHAGRLDAAVEPVIFSDMEVKIT 120  
 Db 100 RKAAGGHHYVOMAFMKLAALGTYVYDHLPLODWAHAGRLDAAVEPVIFSDMEVKIT 159  
 Qy 121 WGADPAACDIISGLPVASARRBREILGPADPEFGQWRLLAPITAVSQQRGLGICIT 180  
 Db 160 WGADPAACDIISGLPVASARRBREILGPADPEFGQWRLLAPITAVSQQRGLGICIT 219  
 Qy 181 SLTGRDKNQVEGVQVSTATOSFLATCNGCWTWFGASSKTLAGPKGPIQMTNVD 240  
 Db 220 SLTGRDKNQVEGVQVSTATOSFLATCNGCWTWFGASSKTLAGPKGPIQMTNVD 279  
 Qy 241 QDLVGWQAPPGRGARSMTPTCGSSDLYLVTRHADVIPRRGDSRGSLSPPRVSYLKGS 300  
 Db 280 QDLVGWQAPPGRGARSMTPTCGSSDLYLVTRHADVIPRRGDSRGSLSPPRVSYLKGS 339  
 Qy 301 GGPLCPSPGHAVGIFRAAVCTRGAKAVDPIVEMETMR 341  
 Db 340 GGPLCPSPGHAVGIFRAAVCTRGAKAVDPIVEMETMR 380

RESULT 6  
 US-10-650-585-12  
 ; Sequence 12, Application US/10017736  
 ; Publication No. US20020192640A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Boehringer Ingelheim (Canada) Ltd.  
 ; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease  
 ; FILE REFERENCE: 13/082

Query Match 100.0%; Score 1778; DB 13; Length 393;  
 Best Local Similarity 100.0%; Pred. No. 3.4e-174; Matches 341; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VRGGDAAILLTCAVHPELIFDITKLLIAIFGPLMVLQAGITKPYFVRAQGLIRACMV 60  
 Db 53 VRGGDAAILLTCAVHPELIFDITKLLIAIFGPLMVLQAGITKPYFVRAQGLIRACMV 112  
 Qy 61 RKAAGGHHYVOMAFMKLAALGTYVYDHLPLODWAHAGRLDAAVEPVIFSDMEVKIT 120

QY 113 RKAAGHYYVQAFMILIAALTTGTYVVDLTHLTLQDWAHAGRLDAVAYEPVIFSDMEVKIT 172

QY 121 WGADTAACGDIISGLPVSRARGRELLGPADNFGQWRLAPITAYSQTRGLGCIT 180

Db 173 WGADTAACGDIISGLPVSRARGRELLGPADNFGQWRLAPITAYSQTRGLGCIT 232

QY 181 SLTGDRDNQVEGEVQVSTATOSFLATCNGVCMVFHGAGSKILAGPKPITOYNTND 240

Db 233 SLTGDRDNQVEGEVQVSTATOSFLATCNGVCMVFHGAGSKILAGPKPITOYNTND 292

QY 241 QDLVGWQAPPGRASMTPTCGSSDLYLVTRHADVIPVRRGDSRSLSSPRVSYLGSS 300

Db 293 QDLVGWQAPPGRASMTPTCGSSDLYLVTRHADVIPVRRGDSRSLSSPRVSYLGSS 352

QY 301 GGPLICPSGHAVGIFRAAVCTRGVAKAVDPFIPVEMETMR 341

Db 353 GGPLICPSGHAVGIFRAAVCTRGVAKAVDPFIPVEMETMR 393

RESULT 8

; Sequence 11, Application US/10650585

; Publication No. US20040077066A1

; GENERAL INFORMATION:

; APPLICANT: Boehringer Ingelheim (Canada) Ltd.

; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease

; FILE REFERENCE: 13/082

; CURRENT APPLICATION NUMBER: US/10/017,736

; CURRENT FILING DATE: 2001-12-14

; PRIOR APPLICATION NUMBER: 60/256,031

; PRIOR FILING DATE: 2000-12-15

; NUMBER OF SEQ ID NOS: 21

; SEQ ID NO: 2

; LENGTH: 409

; TYPE: PRT

; ORGANISM: HCV

US-10-017-736-2

Query Match 100.0%; Score 1778; DB 16; Length 409;

Best Local Similarity 100.0%; Pred. No. 3\_6e-174; Length 409;

Matches 341; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VRGGDRAILLTCAVHPPELFIDTCKLIAIFGPIMVLQAGITKPYFVRAGQLIRACMLV 60

Db 53 VRGGDRAILLTCAVHPPELFIDTCKLIAIFGPIMVLQAGITKPYFVRAGQLIRACMLV 112

QY 61 RKAAGHYYVQAFMILIAALTTGTYVVDLTHLTLQDWAHAGRLDAVAYEPVIFSDMEVKIT 120

Db 117 RKAAGHYYVQAFMILIAALTTGTYVVDLTHLTLQDWAHAGRLDAVAYEPVIFSDMEVKIT 176

QY 121 WGADTAACGDIISGLPVSRARGRELLGPADNFGQWRLAPITAYSQTRGLGCIT 180

Db 177 WGADTAACGDIISGLPVSRARGRELLGPADNFGQWRLAPITAYSQTRGLGCIT 236

QY 181 SLTGDRDNQVEGEVQVSTATOSFLATCNGVCMVFHGAGSKILAGPKPITOYNTND 240

Db 237 SLTGDRDNQVEGEVQVSTATOSFLATCNGVCMVFHGAGSKILAGPKPITOYNTND 296

QY 241 QDLVGWQAPPGRASMTPTCGSSDLYLVTRHADVIPVRRGDSRSLSSPRVSYLGSS 300

Db 297 QDLVGWQAPPGRASMTPTCGSSDLYLVTRHADVIPVRRGDSRSLSSPRVSYLGSS 356

QY 301 GGPLICPSGHAVGIFRAAVCTRGVAKAVDPFIPVEMETMR 341

Db 357 GGPLICPSGHAVGIFRAAVCTRGVAKAVDPFIPVEMETMR 397

RESULT 9

; Sequence 2, Application US/10650585

; Publication No. US20040077066A1

; GENERAL INFORMATION:

; APPLICANT: Boehringer Ingelheim (Canada) Ltd.

; TITLE OF INVENTION: Purified Active HCV NS2/3 Protease

; FILE REFERENCE: 13/082

; CURRENT APPLICATION NUMBER: US/10/650,585

; CURRENT FILING DATE: 2003-09-28

; PRIOR APPLICATION NUMBER: US/10/017,736

; PRIOR FILING DATE: 2001-12-14

; NUMBER OF SEQ ID NOS: 21

; SEQ ID NO: 2

; LENGTH: 409

; TYPE: PRT

; ORGANISM: HCV

US-10-017-736-2

Query Match 100.0%; Score 1778; DB 16; Length 409;

Best Local Similarity 100.0%; Pred. No. 3\_6e-174; Length 409;

Matches 341; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VRGGDRAILLTCAVHPPELFIDTCKLIAIFGPIMVLQAGITKPYFVRAGQLIRACMLV 60

RESULT 11

Db 57 VRRGGRAILITTCAVHPPELIFDITKLLAIFGPLVNLQIGITKPYFVRAGLIRACMV 116  
 QY 61 RKAAGGHIVVQMAFMKLAALTTGTYVVDHPLQLQWAHAGERDLAVAVEPVIFPSMEVKIT 120  
 Db 117 RKAAGGHIVVQMAFMKLAALTTGTYVVDHPLQLQWAHAGERDLAVAVEPVIFPSMEVKIT 176  
 QY 121 WGADTAACGGLISGIPVSARGRETLIGPAPNFQGQWILLAPITAYTOSQTGRGLGCIT 180  
 Db 177 WGADTAACGGLISGIPVSARSGRELLIGPAPNFQGQWILLAPITAYTOSQTGRGLGCIT 236  
 QY 181 SLTGRDKNQVEGEVGVVSTATOSFLATCNGCWTYHGAGSKLAGPKGP-TOMYNTWD 240  
 Db 237 SLTGRDKNQVEGEVGVVSTATOSFLATCNGCWTYHGAGSKLAGPKGP-TOMYNTWD 296  
 QY 241 QDLVWQAPPGRSMPTCGSSDLYLVTYRHDVTPVRRGDRGSLSPPRPSYLGSS 300  
 Db 297 QDLVWQAPPGRSMPTCGSSDLYLVTYRHDVTPVRRGDRGSLSPPRPSYLGSS 356  
 QY 301 GGPLICPSPSHAVGIFRAAVCTRGVAKAVDPIPVESMETMR 341  
 Db 357 GGPLICPSPSHAVGIFRAAVCTRGVAKAVDPIPVESMETMR 397

RESULT 12

Db 297 QDLVWQAPPGRSMPTCGSSDLYLVTYRHDVTPVRRGDRGSLSPPRPSYLGSS 356  
 QY 301 GGPLICPSPSHAVGIFRAAVCTRGVAKAVDPIPVESMETMR 341  
 Db 357 GGPLICPSPSHAVGIFRAAVCTRGVAKAVDPIPVESMETMR 397

RESULT 13

US-10-029-907-3

; Sequence 3, Application US/10029907  
; Publication No. US20020142350A1  
; GENERAL INFORMATION:  
; APPLICANT: BOHRINGER INGELHEIM (CANADA) LTD.  
; TITLE OF INVENTION: SELF REPLICATING RNA MOLECULE FROM  
; FILE REFERENCE: 13/083  
; CURRENT APPLICATION NUMBER: US/10/029, 907  
; PRIORITY APPLICATION NUMBER: 60/257, 857  
; PRIORITY FILING DATE: 2000-12-22  
; NUMBER OF SEQ ID NOS: 25  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 2201  
; TYPE: PRT  
; ORGANISM: HCV  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: 882  
; OTHER INFORMATION: Xaa is Lys or Arg  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: 1489  
; OTHER INFORMATION: Xaa is Leu

US-10-309-561-3

Query Match 95.7%; Score 1702; DB 14; Length 2201;  
 Best Local Similarity 94.1%; Pred. No. 2.6e-165; Matches 321; Conservative 11; Mismatches 9; Indels 0; Gaps 0;

QY 1 VRGGRAILITTCAVHPPELIFDITKLLAIFGPLVNLQIGITKPYFVRAGLIRACMV 60  
 Db 57 VRRGGRAILITTCAVHPPELIFDITKLLAIFGPLVNLQIGITKPYFVRAGLIRACMV 116  
 QY 61 RKAAGGHIVVQMAFMKLAALTTGTYVVDHPLQLQWAHAGERDLAVAVEPVIFPSMEVKIT 120  
 Db 117 RKAAGGHIVVQMAFMKLAALTTGTYVVDHPLQLQWAHAGERDLAVAVEPVIFPSMEVKIT 176  
 QY 121 WGADTAACGGLISGIPVSARGRETLIGPAPNFQGQWILLAPITAYTOSQTGRGLGCIT 180  
 Db 177 WGADTAACGGLISGIPVSARSGRELLIGPAPNFQGQWILLAPITAYTOSQTGRGLGCIT 236  
 QY 181 SLTGRDKNQVEGEVGVVSTATOSFLATCNGCWTYHGAGSKLAGPKGP-TOMYNTWD 240  
 Db 237 SLTGRDKNQVEGEVGVVSTATOSFLATCNGCWTYHGAGSKLAGPKGP-TOMYNTWD 296  
 QY 241 QDLVWQAPPGRSMPTCGSSDLYLVTYRHDVTPVRRGDRGSLSPPRPSYLGSS 300  
 Db 297 QDLVWQAPPGRSMPTCGSSDLYLVTYRHDVTPVRRGDRGSLSPPRPSYLGSS 356  
 QY 301 GGPLICPSPSHAVGIFRAAVCTRGVAKAVDPIPVESMETMR 341  
 Db 357 GGPLICPSPSHAVGIFRAAVCTRGVAKAVDPIPVESMETMR 397

RESULT 13

US-10-467-000-1

; Sequence 1, Application US/10467000  
; Publication No. US20040067486A1  
; GENERAL INFORMATION:

APPLICANT: De Francesco, Raffaele  
APPLICANT: Migliaccio, Giovanni  
APPLICANT: Pionesi, Giacomo

TITLE OF INVENTION: HEPATITIS C VIRUS REPLICONS AND REPLICON

FILE REFERENCE: JTR003P

CURRENT APPLICATION NUMBER: US/10/467,000

PRIOR APPLICATION NUMBER: PCT/EP02/00526

PRIOR FILING DATE: 2002-01-16

PRIOR APPLICATION NUMBER: 60/263,479

PRIOR FILING DATE: 2001-01-23

NUMBER OF SEQ ID NOS: 13

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 1

LENGTH: 3101

TYPE: PRT

ORGANISM: Con 1 HCV isolate nucleic acid

US-10-467-000-1

Query Match 95.7%; Score 1702; DB 12; Length 3101;  
Best Local Similarity 94.1%; Pred. No. 4.6e-165; Indels 0; Gaps 0;  
Matches 320; Conservative 12; Mismatches 8; Indels 0; Gaps 0;

Qy 1 VRGGDAILITCAVHPELIFDTIKLLAIFGPMVHQAGITKPYFVRAGQIYRACMLVR 60  
Db 866 VRGGRDAVILTCIAHPELIFDTIKLLAIFGPMVHQAGITKPYFVRAGQIYRACMLVR 925

Qy 61 RKGAGHIVOMAFMKLGALTGTIVYVHLPDQWHAHGLDLAVEPVFSMEVKIT 120  
Db 926 RKGAGHIVOMAFMKLGALTGTIVYVHLPDQWHAHGLDLAVEPVFSMEVKIT 985

Qy 121 WGADTAACGDIISGLPVRSARGRETEINGPADNFEQGMRLAPITAYSQOTRGILCIT 180  
Db 986 WGADTAACGDIISGLPVRSARGRETEINGPADNFEQGMRLAPITAYSQOTRGILCIT 1045

Qy 181 SLTGRDKNQVEGVQVSTATOSFLATCNGVQVTFHGRSKLALGPKPTQMYNVD 240  
Db 1046 SLTGRDRNQVEGVQVSTATOSFLATCNGVQVTFHGRSKLALGPKPTQMYNVD 1105

Qy 241 QLVGKQAPARASMPCTCSDLYLTHADIVPVRRGDGSLSLSPRPSYLIKSS 300  
Db 1106 QDLVGWQAPGARSLSLTPCTCSDLYLTHADIVPVRRGDGSLSLSPRPSYLIKSS 1165

Qy 301 GPPICSGHAGVIFRAAVCORGVAKAVDIPVSMETMR 341

Db 1166 GPPICSGHAGVIFRAAVCORGVAKAVDIPVSMETMR 1206

#### RESULT 15

US-10-191-966-4  
Sequence 4, Application US/10/191966

Publication No. US20030175692A1

GENERAL INFORMATION:

APPLICANT: Potts, Karen E.

APPLICANT: Jackson, Roberta L.

APPLICANT: Patick, Amy K.

TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT

TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE

FILE REFERENCE: 0125-005A

CURRENT APPLICATION NUMBER: US/10/191-966

CURRENT FILING DATE: 2002-07-10

PRIOR APPLICATION NUMBER: US/09/263,933

PRIOR FILING DATE: 1999-03-09

PRIOR APPLICATION NUMBER: 09/129,611

PRIOR FILING DATE: 1998-08-05

NUMBER OF SEQ ID NOS: 33

SEQ ID NO 4

LENGTH: 1692

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE: OTHER INFORMATION: :

US-10-191-966-4

Query Match 95.5%; Score 1698; DB 14; Length 1692;

Best Local Similarity 94.1%; Pred. No. 4.6e-165; Indels 0; Gaps 0;

Matches 320; Conservative 12; Mismatches 8; Indels 0; Gaps 0;

Qy 2 RGGDAILITCAVHPELIFDTIKLLAIFGPMVHQAGITKPYFVRAGQIYRACMLVR 61  
Db 146 RGGDAILITCAVHPELIFDTIKLLAIFGPMVHQAGITKPYFVRAGQIYRACMLVR 205  
Qy 62 KAAGHIVOMAFMKLGALTGTIVYVHLPDQWHAHGLDLAVEPVFSMEVKIT 122  
Db 206 KVAGHIVOMAFMKLGALTGTIVYVHLPDQWHAHGLDLAVEPVFSMEVKIT 265  
Qy 122 GADTAACGDIISGLPVRSARGRETEINGPADNFEQGMRLAPITAYSQOTRGILCIT 181  
Db 266 GADTAACGDIISGLPVRSARGRETEINGPADNFEQGMRLAPITAYSQOTRGILCIT 325  
Db 326 LTGRDKNQVEGVQVSTATOSFLATCNGVQVTFHGRSKLALGPKPTQMYNVD 385

Qy 242 DLVGRQAPPGARSMTPCCSSDLYLTHADIVPVRRGDGSLSLSPRPSYLIKSS 301  
Db 386 DLVGRQAPPGARSLSLTPCTCGSSDLYLTHADIVPVRRGDGSLSLSPRPSYLIKSS 445

Qy 302 GPLICPSGHAGVIFRAAVCORGVAKAVDIPVSMETMR 341  
Db 446 GPLICPSGHAGVIFRAAVCORGVAKAVDIPVSMETMR 485

RESULT 14

US-09-919-901-4

Sequence 4, Application US/09/19901

Publication No. US20030175692A1

GENERAL INFORMATION:

APPLICANT: Potts, Karen E.

APPLICANT: Jackson, Roberta L.

APPLICANT: Patick, Amy K.

TITLE OF INVENTION: REPORTER GENE SYSTEM FOR USE IN CELL-BASED ASSESSMENT

TITLE OF INVENTION: OF INHIBITORS OF THE HEPATITIS C VIRUS PROTEASE

FILE REFERENCE: 0125-005A

CURRENT APPLICATION NUMBER: US/10/919-901

CURRENT FILING DATE: 2001-08-02

PRIOR APPLICATION NUMBER: 09/263,933

PRIOR FILING DATE: 1999-02-08

PRIOR APPLICATION NUMBER: 09/129,611

PRIOR FILING DATE: 1998-05-05

NUMBER OF SEQ ID NOS: 33

SEQ ID NO 4

LENGTH: 1692

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE: OTHER INFORMATION: :

US-10-191-966-4

Query Match 95.5%; Score 1698; DB 14; Length 1692;

Best Local Similarity 94.1%; Pred. No. 4.6e-165; Indels 0; Gaps 0;

Matches 320; Conservative 12; Mismatches 8; Indels 0; Gaps 0;

Qy 2 RGGDAILITCAVHPELIFDTIKLLAIFGPMVHQAGITKPYFVRAGQIYRACMLVR 61  
Db 146 RGGDAILITCAVHPELIFDTIKLLAIFGPMVHQAGITKPYFVRAGQIYRACMLVR 205  
Qy 62 KAAGHIVOMAFMKLGALTGTIVYVHLPDQWHAHGLDLAVEPVFSMEVKIT 122  
Db 206 KVAGHIVOMAFMKLGALTGTIVYVHLPDQWHAHGLDLAVEPVFSMEVKIT 265  
Qy 122 GADTAACGDIISGLPVRSARGRETEINGPADNFEQGMRLAPITAYSQOTRGILCIT 181

Fri May 7 13:37:05 2004

us-10-650-585-14.rapb

Page 7

Db 266 GADTAACGDTILGLPVSAARRKEIILGPADSLEGRGWRLIAPITAYSOOTRGLGICITS 325  
Qy 182 |TGRKRNQVEGEVQYSTATCPLATCYNCWTFHAGSKTLAGPKPITOYNTDQ 241  
Db 326 |TGRDKNQVEGEVQYSTATCPLATCYNCWTFHAGSKTLAGPKPITOYNTDQ 385  
Qy 242 DLVGMQAPPGRSMPTCGSDLXLYLVRHADVIPVRRGDSRSLSPRPVSYLKSSG 301  
Db 386 DLVGMQAPPGRSLPCTCGSDLXLYLVRHADVIPVRRGDSRSLSPRPVSYLKSSG 445  
Qy 302 GPLLCPSGHAVGIFRAVCIRGVAKADVIPVESMETMR 341  
Db 446 GPLLCPSGHAVGIFRAVCIRGVAKADVIPVESMETMR 485

Search completed: May 6, 2004, 09:43:22  
Job time : 35.8831 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2004, 09:08:45 ; Search time 39.7461 Seconds  
{without alignments}  
2075.771 Million cell updates/sec

Title: US-10-650-585-15  
Perfect score: 1532  
Sequence: I A Q G L I R A C M L V R K A G G H Y V . . . . . R G V A K A V D F I P V E S M E T T R V 292

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:  
1: geneseqp1980s: \*  
2: geneseqp1900s: \*  
3: geneseqp2000s: \*  
4: geneseqp2001s: \*  
5: geneseqp2002s: \*  
6: geneseqp2003as: \*  
7: geneseqp2003bs: \*  
8: geneseqp2004as: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

**SUMMARIES**

| Result No. | Score | Query Length | DB ID | Description |
|------------|-------|--------------|-------|-------------|
| 1          | 1532  | 100.0        | 292   | ABG32188    |
| 2          | 1532  | 100.0        | 303   | ABG32183    |
| 3          | 1532  | 100.0        | 334   | ABG32182    |
| 4          | 1532  | 100.0        | 341   | ABG32187    |
| 5          | 1532  | 100.0        | 352   | ABG32186    |
| 6          | 1532  | 100.0        | 380   | ABG32185    |
| 7          | 1532  | 100.0        | 393   | ABG32184    |
| 8          | 1532  | 100.0        | 409   | ABG32181    |
| 9          | 1523  | 99.4         | 303   | ABG32191    |
| 10         | 1522  | 99.3         | 303   | ABG32189    |
| 11         | 1513  | 98.8         | 301   | ABG32190    |
| 12         | 1481  | 96.7         | 768   | AAR0223     |
| 13         | 1476  | 96.3         | 3010  | AAE20477    |
| 14         | 1473  | 96.1         | 2201  | ABG30601    |
| 15         | 1473  | 96.1         | 2201  | ABG30591    |
| 16         | 1473  | 96.1         | 2201  | ABG30600    |
| 17         | 1473  | 96.1         | 2201  | ABG30581    |
| 18         | 1473  | 96.1         | 2201  | ABG30593    |
| 19         | 1473  | 96.1         | 2201  | ABG30582    |
| 20         | 1473  | 96.1         | 2201  | ABG30580    |
| 21         | 1473  | 96.1         | 2201  | ABG30587    |
| 22         | 1473  | 96.1         | 2201  | ABG30599    |
| 23         | 1473  | 96.1         | 2201  | ABG30594    |
| 24         | 1473  | 96.1         | 2201  | ABG30598    |
| 25         | 1473  | 96.1         | 2201  | ABG30595    |

**ALIGNMENTS**

| RESULT | ID       | NAME  | TYPE | DESCRIPTION   |
|--------|----------|---|------|---|
| 1      | ABG32188 | standard; protein; 292 AA.  | XX   | ABG32188;   |
| 2      | ABG32188 | (first entry)   | DT   | 05-NOV-2002   |
| 3      | ABG32188 | HCV protease NS2/3 truncation mutant 915-1206.  | XX   | HCV; enzyme; protease; NS2/3 (915-1206); hepatitis C virus infection; chronic liver disease; cirrhosis; end-stage liver disease; viricide; hepatotropic; anti-inflammatory; lauryldiethylamine oxide; LDAO; chaotropic agent; mutant; mutein.   |
| 4      | ABG32188 | OS Synthetic.   | XX   | OS Hepatitis C virus.   |
| 5      | ABG32188 | OS Synthetic.   | XX   | OS  |
| 6      | ABG32188 | WO200248375-A2.   | XX   | WO200248375-A2.   |
| 7      | ABG32188 | PD 20-JUN-2002.   | XX   | PD  |
| 8      | ABG32188 | PR 13-DEC-2001; 2001WO-CA001795.  | XX   | PR  |
| 9      | ABG32188 | PR 15-DEC-2000; 2000US-0256031P.  | XX   | PR  |
| 10     | ABG32188 | (BÖHRL ) BOEHRINGER INGELHEIM CANADA LTD.   | PA   | (BÖHRL ) BOEHRINGER INGELHEIM CANADA LTD.   |
| 11     | ABG32188 | PI Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;  | XX   | PI Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;  |
| 12     | ABG32188 | XX DR WPI; 2002-595511/64.  | XX   | DR WPI; 2002-595511/64.   |
| 13     | ABG32188 | PT Novel polypeptide for screening inhibitors of non-structural Proteases useful as therapeutic agents against hepatitis C virus, comprises full length non-structural protease, or its truncation.   | XX   | PT Novel polypeptide for screening inhibitors of non-structural Proteases useful as therapeutic agents against hepatitis C virus, comprises full length non-structural protease, or its truncation.   |
| 14     | ABG32188 | PT claim 41; Page 63; 67pp; English.  | XX   | PT claim 41; Page 63; 67pp; English.  |
| 15     | ABG32188 | CC The invention relates to an isolated polypeptide consisting of a full-length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal residue amino acid 810 to 906, or having a minimal amino acid sequence from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length NS2/3 protease. Also included are (1) a composition (C) comprising an isolated HCV NS2/3 protease selected from the full length NS2/3 protease, or its truncation or a mutated sequence, where the protease is in a solution comprising a sufficient concentration of lauryldiethylamine oxide (LDAO) to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitor peptide | CC   | CC The invention relates to an isolated polypeptide consisting of a full-length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal residue amino acid 810 to 906, or having a minimal amino acid sequence from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length NS2/3 protease. Also included are (1) a composition (C) comprising an isolated HCV NS2/3 protease selected from the full length NS2/3 protease, or its truncation or a mutated sequence, where the protease is in a solution comprising a sufficient concentration of lauryldiethylamine oxide (LDAO) to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitor peptide |
| 16     | ABG32188 | CC  | CC   | CC  |
| 17     | ABG32188 | CC  | CC   | CC  |
| 18     | ABG32188 | CC  | CC   | CC  |
| 19     | ABG32188 | CC  | CC   | CC  |
| 20     | ABG32188 | CC  | CC   | CC  |
| 21     | ABG32188 | CC  | CC   | CC  |
| 22     | ABG32188 | CC  | CC   | CC  |
| 23     | ABG32188 | CC  | CC   | CC  |
| 24     | ABG32188 | CC  | CC   | CC  |
| 25     | ABG32188 | CC  | CC   | CC  |

CC appearing as ABC32198; (3) producing (M1) a refolded, inactive HCV NS2/3 protease, involving isolating the protease in the presence of a chaotropic agent, refolding the isolated protease by contacting it with a reducing agent, and LDAO in the presence of reduced concentration of the chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3 protease, involving diluting refolded inactive NS2/3 protease in a medium -protease; (5) measuring (M3) the auto-cleavage activity of NS2/3 protease, involving incubating the active NS2/3 protease produced by M2 for sufficient time to induce auto-cleavage of NS2/3 protease and produce cleavage products or their fragments, and measuring the presence or absence of uncleaved NS2/3 protease, cleavage products or their fragments ; and (6) screening a potential inhibitor of auto-cleavage activity of an active NS2/3 protease, involving carrying out M3 in the presence of, or absence of the potential inhibitor, comparing the amount of uncleaved NS2/3 protease, cleavage products or their fragments. The protease is useful for detailed biochemical characterisation of the enzymes and in the development of in vitro assays for screening novel inhibitors of NS2/3 protease which are useful as therapeutic agents against HCV infection (which causes chronic liver disease, cirrhosis and end-stage liver disease). M1 is useful for high level production of protease. The present sequence represents the NS2/3 truncation mutant 915-1206 (numbered relative to the full length NS2/3 protein)

XX SQ Sequence 292 AA;

Query Match 100.0%; Score 1532; DB 5; Length 292;  
Best Local Similarity 100.0%; Pred. No. 1-2e-140; Indels 0; Gaps 0;  
Matches 292; Conservative 0; Mismatches 0; Sequence 292 AA;

QY 1 AQLIRACMLVRKAAGGHYQMAFKLALITGTYVYDHTPLQDWAHGLRDLAVAEPV 60  
Db 1 AQLIRACMLVRKAAGGHYQMAFKLALITGTYVYDHTPLQDWAHGLRDLAVAEPV 60  
QY 61 IFSDMVKITWGAATACCDIISGLPVSAARRGRBILIGPADNFEQGWRILLAPITAYQ 120  
Db 61 IFSDMVKITWGAATACCDIISGLPVSAARRGRBILIGPADNFEQGWRILLAPITAYQ 120  
QY 121 QTRGLIGCILTSLGDRDKNQVEGEVQVNSTATOSFLATCNGVWTWFGAGSKTLGPK 180  
Db 121 QTRGLIGCILTSLGDRDKNQVEGEVQVNSTATOSFLATCNGVWTWFGAGSKTLGPK 180  
QY 181 GPITQMTNTDQDLYQWQAPGARSMPTCTCGSSDLYLVRHADVIPRRGDSRGSLJS 240  
Db 181 GPITQMTNTDQDLYQWQAPGARSMPTCTCGSSDLYLVRHADVIPRRGDSRGSLJS 240  
QY 241 PRPVSYLKGSGGPILCPSPSHAVGIFRAVCTREVAKAVTFIVESMETMR 292  
Db 241 PRPVSYLKGSGGPILCPSPSHAVGIFRAVCTREVAKAVTFIVESMETMR 292

XX SQ Sequence 303 AA;

RESULT 2  
ABC32183 standard; protein: 303 AA.  
AC ABG32183;  
XX DT 05-NOV-2002 (first entry)  
XX DE HCV protease NS2/3 truncation mutant 904-1206.  
XX KW chronic liver disease; cirrhosis; end-stage liver disease; virucide; hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO; chaotic agent; mutant; muttein.  
OS Hepatitis C virus.  
OS Synthetic.  
XX WO2002A8375-A2.  
XX PN 20-JUN-2002.

XX DR PT Thibeault D, Lamare D, Maurice R, Pilote L, Pause A;  
XX PA (BOEH ) BOHRINGER INGELHEIM CANADA LTD.  
XX PI Thibeault D, Lamare D, Maurice R, Pilote L, Pause A;  
XX PT WPI; 2002-599511/64.

PS Claim 39; Page 58-59; 67pp; English.

CC Novel polypeptide for screening inhibitors of non-structural proteases useful as therapeutic agents against hepatitis C virus, comprises full length non-structural protease, or its truncation.

XX PT Novel polypeptide for screening inhibitors of non-structural proteases useful as therapeutic agents against hepatitis C virus, comprises full length non-structural protease, or its truncation.

XX CC The invention relates to an isolated polypeptide consisting of a full-length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred to also as NS2/3 (910-1206)), or its truncation, having as its N-terminal residue amino acid 810 to 906, or having a minimal amino acid sequence from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length NS2/3 protease. Also included are (1) a composition (C) comprising an isolated HCV NS2/3 protease selected from full length NS2/3 protease, or its truncation or a mutated sequence, where the protease is in a solution comprising a sufficient concentration of lauryldiethylamine oxide (LDAO) to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide appearing as ABC32198; (3) producing (M1) a refolded, inactive HCV NS2/3 protease, involving isolating the protease in the presence of a chaotropic agent, refolding the isolated protease by contacting it with a reducing agent, and LDAO in the presence of reduced concentration of the chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3 protease, involving diluting refolded inactive NS2/3 protease in a medium containing an activation detergent to induce auto-cleavage of the NS2/3 protease; (5) measuring (M3) the auto-cleavage activity of NS2/3 protease, involving incubating the active NS2/3 protease produced by M2 for sufficient time to induce auto-cleavage of NS2/3 Protease and produce cleavage products or their fragments, and measuring the presence or absence of uncleaved NS2/3 protease, cleavage products or their fragments ; and (6) screening a potential inhibitor of auto-cleavage activity of an active NS2/3 protease, involving carrying out M3 in the presence of, or absence of the potential inhibitor, comparing the amount of uncleaved NS2/3 protease, cleavage products or their fragments. The protease is useful for detailed biochemical characterisation of the enzymes and in the development of in vitro assays for screening novel inhibitors of NS2/3 protease which are useful as therapeutic agents against HCV infection (which causes chronic liver disease, cirrhosis and end-stage liver disease). M1 is useful for high level production of protease. The present sequence represents the NS2/3 truncation mutant 904-1206 (numbered relative to the full length NS2/3 protein)

XX SQ Sequence 303 AA;

Query Match 100.0%; Score 1532; DB 5; Length 303;  
Best Local Similarity 100.0%; Pred. No. 1-2e-140; Indels 0; Gaps 0;  
Matches 292; Conservative 0; Mismatches 0; Sequence 303 AA;

QY 1 AQLIRACMLVRKAAGGHYQMAFKLALITGTYVYDHTPLQDWAHGLRDLAVAEPV 60  
Db 1 AQLIRACMLVRKAAGGHYQMAFKLALITGTYVYDHTPLQDWAHGLRDLAVAEPV 71  
QY 61 IFSDMVKITWGAATACCDIISGLPVSAARRGRBILIGPADNFEQGWRILLAPITAYQ 120  
Db 61 IFSDMVKITWGAATACCDIISGLPVSAARRGRBILIGPADNFEQGWRILLAPITAYQ 131  
QY 121 QTRGLIGCILTSLGDRDKNQVEGEVQVNSTATOSFLATCNGVWTWFGAGSKTLGPK 180  
Db 121 QTRGLIGCILTSLGDRDKNQVEGEVQVNSTATOSFLATCNGVWTWFGAGSKTLGPK 191  
QY 181 GPITQMTNTDQDLYQWQAPGARSMPTCTCGSSDLYLVRHADVIPRRGDSRGSLJS 240  
Db 181 GPITQMTNTDQDLYQWQAPGARSMPTCTCGSSDLYLVRHADVIPRRGDSRGSLJS 251  
QY 241 PRPVSYLKGSGGPILCPSPSHAVGIFRAVCTREVAKAVTFIVESMETMR 292

PF 13-DEC-2001; 2001WO-CA001796.  
XX PR 15-DEC-2000; 2000US-0256031P.

CC (BOEH ) BOHRINGER INGELHEIM CANADA LTD.

CC Thibeault D, Lamare D, Maurice R, Pilote L, Pause A;

XX DR WPI; 2002-599511/64.

PS Claim 39; Page 58-59; 67pp; English.

CC Novel polypeptide for screening inhibitors of non-structural proteases useful as therapeutic agents against hepatitis C virus, comprises full length non-structural protease, or its truncation.

XX PT Novel polypeptide for screening inhibitors of non-structural proteases useful as therapeutic agents against hepatitis C virus, comprises full length non-structural protease, or its truncation.

XX CC The invention relates to an isolated polypeptide consisting of a full-length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred to also as NS2/3 (910-1206)), or its truncation, having as its N-terminal residue amino acid 810 to 906, or having a minimal amino acid sequence from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length NS2/3 protease. Also included are (1) a composition (C) comprising an isolated HCV NS2/3 protease selected from full length NS2/3 protease, or its truncation or a mutated sequence, where the protease is in a solution comprising a sufficient concentration of lauryldiethylamine oxide (LDAO) to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide appearing as ABC32198; (3) producing (M1) a refolded, inactive HCV NS2/3 protease, involving isolating the protease in the presence of a chaotropic agent, refolding the isolated protease by contacting it with a reducing agent, and LDAO in the presence of reduced concentration of the chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3 protease, involving diluting refolded inactive NS2/3 protease in a medium containing an activation detergent to induce auto-cleavage of the NS2/3 protease; (5) measuring (M3) the auto-cleavage activity of NS2/3 protease, involving incubating the active NS2/3 protease produced by M2 for sufficient time to induce auto-cleavage of NS2/3 Protease and produce cleavage products or their fragments, and measuring the presence or absence of uncleaved NS2/3 protease, cleavage products or their fragments ; and (6) screening a potential inhibitor of auto-cleavage activity of an active NS2/3 protease, involving carrying out M3 in the presence of, or absence of the potential inhibitor, comparing the amount of uncleaved NS2/3 protease, cleavage products or their fragments. The protease is useful for detailed biochemical characterisation of the enzymes and in the development of in vitro assays for screening novel inhibitors of NS2/3 protease which are useful as therapeutic agents against HCV infection (which causes chronic liver disease, cirrhosis and end-stage liver disease). M1 is useful for high level production of protease. The present sequence represents the NS2/3 truncation mutant 904-1206 (numbered relative to the full length NS2/3 protein)

PF 13-DEC-2001; 2001WO-CA001796.  
XX PR 15-DEC-2000; 2000US-0256031P.

CC (BOEH ) BOHRINGER INGELHEIM CANADA LTD.

CC Thibeault D, Lamare D, Maurice R, Pilote L, Pause A;

XX DR WPI; 2002-599511/64.

PS Claim 39; Page 58-59; 67pp; English.

CC Novel polypeptide for screening inhibitors of non-structural proteases useful as therapeutic agents against hepatitis C virus, comprises full length non-structural protease, or its truncation.

XX PT Novel polypeptide for screening inhibitors of non-structural proteases useful as therapeutic agents against hepatitis C virus, comprises full length non-structural protease, or its truncation.

XX CC The invention relates to an isolated polypeptide consisting of a full-length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred to also as NS2/3 (910-1206)), or its truncation, having as its N-terminal residue amino acid 810 to 906, or having a minimal amino acid sequence from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length NS2/3 protease. Also included are (1) a composition (C) comprising an isolated HCV NS2/3 protease selected from full length NS2/3 protease, or its truncation or a mutated sequence, where the protease is in a solution comprising a sufficient concentration of lauryldiethylamine oxide (LDAO) to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide appearing as ABC32198; (3) producing (M1) a refolded, inactive HCV NS2/3 protease, involving isolating the protease in the presence of a chaotropic agent, refolding the isolated protease by contacting it with a reducing agent, and LDAO in the presence of reduced concentration of the chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3 protease, involving diluting refolded inactive NS2/3 protease in a medium containing an activation detergent to induce auto-cleavage of the NS2/3 protease; (5) measuring (M3) the auto-cleavage activity of NS2/3 protease, involving incubating the active NS2/3 protease produced by M2 for sufficient time to induce auto-cleavage of NS2/3 Protease and produce cleavage products or their fragments, and measuring the presence or absence of uncleaved NS2/3 protease, cleavage products or their fragments ; and (6) screening a potential inhibitor of auto-cleavage activity of an active NS2/3 protease, involving carrying out M3 in the presence of, or absence of the potential inhibitor, comparing the amount of uncleaved NS2/3 protease, cleavage products or their fragments. The protease is useful for detailed biochemical characterisation of the enzymes and in the development of in vitro assays for screening novel inhibitors of NS2/3 protease which are useful as therapeutic agents against HCV infection (which causes chronic liver disease, cirrhosis and end-stage liver disease). M1 is useful for high level production of protease. The present sequence represents the NS2/3 truncation mutant 904-1206 (numbered relative to the full length NS2/3 protein)

Db 252 PRPVSYLKGSQGGPLCPSGHAVGIFRAVCTRGVAKAVDFEVESMETMR 303  
 RESULT 3  
 ABG32182 ABG32182 standard; protein; 334 AA.  
 ID ABG32182;  
 AC ABG32182;  
 XX DT 05-NOV-2002 (first entry)  
 XX DE HCV protease NS2/3 truncation 4K-6H (904-1206)st-4K.  
 XX KW HCV; enzyme; protease; NS2/3 (866-1206); hepatitis C virus infection; chronic liver disease; cirrhosis; end-stage liver disease; virucide; hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO; chaotropic agent; 4K-6H (904-1206)st-4K; mutant; mutein.  
 XX OS Hepatitis C virus.  
 XX FH Key Location/Qualifiers  
 FT Peptide 1..15 /note= "4-Lys/His tag"  
 FT Protein 16..302 /note= "Truncated NS2/3 protease"  
 FT Peptide 319..334 /note= "Streptavidin/4-Lys tag"  
 XX PN WO200248375-A2.  
 XX PD 20-JUN-2002.  
 PF 13-DEC-2001; 2001WO-CA001796.  
 XX PR 15-DEC-2000; 2000US-0256031P.  
 XX PA (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  
 XX PI Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;  
 XX DR WPI; 2002-599511/64.  
 XX DR N-PSDB; ABK90407.  
 XX PT Novel polypeptide for screening inhibitors of non-structural proteases  
 PT useful as therapeutic agents against hepatitis C virus, comprises full  
 PT length non-structural protease, or its truncation.  
 PS Claim 39; Fig 9B; 67pp; English.  
 PT The invention relates to an isolated polypeptide consisting of a full-  
 CC length HCV (hepatitis C virus) non-structural protease (referred  
 CC to also as NS2/3 (866-1206)), or its truncation, having as its N-terminal  
 CC amino acid 866 to 905, or having a minimal amino acid sequence  
 CC from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length  
 CC NS2/3 protease. Also included are (1) a composition (C) comprising an  
 CC isolated HCV NS2/3 protease selected from full length NS2/3 protease, or  
 CC its truncation or a mutated sequence, where the protease is in a solution  
 CC comprising a sufficient concentration of lauryldiethylamine oxide (LDAO),  
 CC to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitor peptide  
 CC appearing as ABG32182; (3) producing (M1) a refolded, inactive HCV NS2/3  
 CC protease, involving isolating the protease in the presence of a  
 CC chaotropic agent, refolding the isolated protease by contacting it with a  
 CC reducing agent, and LDAO in the presence of reduced concentration of the  
 CC chaoticropic agent or a polar additive; (4) producing (M2) an active NS2/3  
 CC protease, involving diuting refolded inactive NS2/3 protease in a medium  
 CC containing an activation detergent to induce auto-cleavage of the NS2/3  
 CC protease; (5) measuring (M3) the auto-cleavage activity of NS2/3 protease produced by M2  
 CC for sufficient time to induce auto-cleavage of NS2/3 protease and produce  
 CC cleavage products or their fragments, and measuring the presence or  
 CC absence of uncleaved NS2/3 protease, cleavage products or their fragments  
 CC ; and (6) screening a potential inhibitor of auto-cleavage activity of an  
 CC active NS2/3 protease, involving carrying out M3 in the presence of, or  
 CC absence of the potential inhibitor, comparing the amount of uncleaved  
 CC NS2/3 protease, cleavage products or their fragments. The protease is  
 CC useful for detailed biochemical characterisation of the enzymes and in  
 CC the development of in vitro assays for screening novel inhibitors of  
 CC NS2/3 protease which are useful as therapeutic agents against HCV  
 CC infection (which causes chronic liver disease, cirrhosis and end-stage  
 CC liver disease, M1 is useful for high level production of protease. The  
 CC present sequence represents the NS2/3 truncation mutant 4K-6H (904-  
 CC 1206)st-4K comprising a truncated NS2/3 protease with a four Lys/His N  
 CC terminal tag, a C-terminal streptavidin tag and C-terminal four Lys tag  
 XX Sequence 334 AA;  
 Query Match 100 %; Score 152; DB 5; Length 334;  
 Best Local Similarity 100.0%; Pred. No. 1.4e-140;  
 Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 1 AGQLIRACMLVKAAGHHYVQAFMKLALGTGYVIDLTPQLQWAHGJDLAVPE PV 60  
 DB 27 AQGLIRACMLVKAAGHHYVQAFMKLALGTGYVIDLTPQLQWAHGJDLAVPE PV 86  
 OY 61 IFSDMEVKITTGADTAAGCDIISGPVPSARRGRGRETTLGPADNFGQGRWLLAPTTAYSQ 120  
 DB 87 IFSDMEVKITTGADTAAGCDIISGPVPSARRGRGRETTLGPADNFGQGRWLLAPTTAYSQ 146  
 OY 121 QFRGLGCGCITSLTGDKNOYEVEGVVSVTATCAGTCVWVFRAGSKTLAGPK 180  
 DB 147 QTRGLLGCGCITSLTGDKNOYEVEGVVSVTATCAGTCVWVFRAGSKTLAGPK 206  
 OY 181 GRITQWVNIVDVLGVQAPPGARSMTPTCIGSSDLYLVRHADYIPYRERGDSGSLS 240  
 DB 207 GRITQWVNIVDVLGVQAPPGARSMTPTCIGSSDLYLVRHADYIPYRERGDSGSLS 266  
 OY 241 PRPVSYLKGSQGGPLCPSGHAVGIFRAVCTRGVAKAVDFEVESMETMR 292  
 DB 267 PRPVSYLKGSQGGPLCPSGHAVGIFRAVCTRGVAKAVDFEVESMETMR 318  
 RESULT 4  
 ABG32187 ABG32187 standard; protein; 341 AA.  
 ID ABG32187;  
 XX AC ABG32187;  
 XX DT 05-NOV-2002 (first entry)  
 XX DE HCV protease NS2/3 truncation mutant 866-1205.  
 XX KW HCV; enzyme; protease; NS2/3 (866-1206); hepatitis C virus infection; chronic liver disease; cirrhosis; end-stage liver disease; virucide; hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO; chaotropic agent; mutant; mutein.  
 XX OS Hepatitis C virus.  
 XX PN WO200248375-A2.  
 XX PD 20-JUN-2002.  
 XX PF 13-DEC-2001; 2001WO-CA001796.  
 XX PR 15-DEC-2000; 2000US-0256031P.  
 XX PA (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  
 XX PI Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;  
 XX DR WPI; 2002-599511/64.  
 XX PT Novel polypeptide for screening inhibitors of non-structural proteases

useful as therapeutic agents against hepatitis C virus, comprises full length non-structural protease, or its truncation.

XX  
PT  
PS Claim 41; Page 62-63; 67pp; English.

The invention relates to an isolated polypeptide consisting of a full-length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal residue amino acid 810 to 906, or having a minimal amino acid sequence from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length NS2/3 protease. Also included are (1) a composition (C) comprising an isolated HCV NS2/3 protease selected from full length NS2/3 protease, or its truncation or a mutated sequence, where the protease is in a solution comprising a sufficient concentration of lauryldiethylamine oxide (LDAO) to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide appearing as ABG31198; (3) producing (M1) a refolded, inactive HCV NS2/3 protease, involving isolating the protease in the presence of a chaotropic agent, refolding the isolated protease by contacting it with a reducing agent, and LDAO in the presence of reduced concentration of the chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3 protease, involving diluting refolded inactive NS2/3 protease in a medium containing an activation detergent to induce auto-cleavage of the NS2/3 protease; (5) measuring (M3) the auto-cleavage activity of NS2/3 protease, involving incubating the active NS2/3 Protease produced by M2 for sufficient time to induce auto-cleavage of NS2/3 protease and produce cleavage products or their fragments, and measuring the presence or absence of uncleaved NS2/3 protease, cleavage products or their fragments ; and (6) screening a potential inhibitor of auto-cleavage activity of an active NS2/3 protease involving carrying out M3 in the presence of, or absence of, the potential inhibitor, comparing the amount of uncleaved NS2/3 protease, cleavage products or their fragments. The protease is useful for detailed biochemical characterisation of the enzymes and in the development of in vitro assays for screening novel inhibitors of NS2/3 protease which are useful as therapeutic agents against HCV infection (which causes chronic liver disease, cirrhosis and end-stage liver disease). M1 is useful for high level production of protease. The present sequence represents the NS2/3 truncation mutant 866-1206 (numbered relative to the full length NS2/3 protein)

SQ Sequence 341 AA;

Query Match 100.0%; Score 1532; DB 5; Length 341;  
Best Local Similarity 100.0%; Pred. No. 1.5e-140; Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CC 1 AQLIRACMLYRKAGGHYYQMAFKLALLTGTGVYDHITPLQWAHAGIRDIAVAPV 60  
Db 50 AQGLIRACMLYRKAGGHYYQMAFKLALLTGTGVYDHITPLQWAHAGIRDIAVAPV 109  
Qy 61 IPSDEVKITIWGADTAACDDIISLPUVSARRGREGIELLGFDNREGQQWRLLAPIYSQ 120  
Db 110 IFSDMEVKITIWGADTAACDDIISLPUVSARRGRBILIGFADNREGQQWRLLAPIYSQ 169  
Qy 121 QTRGLIGCITSLSLGRDKQVEGVYQVSVTATOSPLATCNGVQWTFHGAASKTLAGPK 180  
Db 170 QTRGLIGCITSLSLGRDKQVEGVYQVSVTATOSPLATCNGVQWTFHGAASKTLAGPK 229  
Qy 181 GPITQMYNTNDQDLYQWPAGSPARMSMPCFGSSDLYLVRHADVIFRRGRSGSILS 240  
Db 230 QPITQMYNTNDQDLYQWPAGSPARMSMPCFGSSDLYLVRHADVIFRRGRSGSILS 289  
Qy 241 PRPVSTIKGGSGGRICLPCPSCHAVGIFRAVCTRVAKAVDFIPVSMITMR 292  
Db 290 PRPVSTIKGGSGGRICLPCPSCHAVGIFRAVCTRVAKAVDFIPVSMITMR 341  
SQ Sequence 352 AA;

Query Match 100.0%; Score 1532; DB 5; Length 352;  
Best Local Similarity 100.0%; Pred. No. 1.5e-140; Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX  
PT  
XX  
AC ABG32186;  
XX DT 05-NOV-2002 (first entry)

XX  
DE HCV protease NS2/3 truncation mutant 855-1206.  
XX HCV; enzyme; protease; NS2/3 (855-1206); hepatitis C virus infection;  
KW chronic liver disease; cirrhosis; end-stage liver disease; viricide;  
KW hepatotoxic; antiinflammatory; lauryldiethylamine oxide; LDAO;  
KW chaotropic agent; mutant; murein.  
XX OS Synthetic.  
XX Hepatitis C virus.

XX WO200248375-A2.  
XX PD 20-JUN-2002.  
XX PR 13-DEC-2001; 2001WO-CA001796.  
XX 15-DEC-2000; 2000US-0256031P.  
XX PA (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  
XX Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;  
XX DR WPI; 2002-599511/64.

XX Novel polypeptide for screening inhibitors of non-structural proteases  
PT useful as therapeutic agents against hepatitis C virus, comprises full  
PR length non-structural protease, or its truncation.  
XX  
Claim 41; Page 61-62; 67pp; English.  
XX The invention relates to an isolated polypeptide consisting of a full-length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred to also as NS2/3 (810-1206)) or its truncation, having as its N-terminal residue amino acid 810 to 906, or having a minimal amino acid sequence from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length NS2/3 protease. Also included are (1) a composition (C) comprising an isolated HCV NS2/3 protease selected from full length NS2/3 protease, or its truncation or a mutated sequence, where the protease is in a solution comprising a sufficient concentration of lauryldiethylamine oxide (LDAO) to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3 protease, involving isolating the protease in the presence of a chaotropic agent, refolding the isolated protease by contacting it with a reducing agent, and LDAO in the presence of reduced concentration of the chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3 protease, involving diluting refolded inactive NS2/3 protease in a medium containing an activation detergent to induce auto-cleavage of the NS2/3 protease; (5) measuring (M3) the auto-cleavage activity of NS2/3 protease, involving incubating the active NS2/3 protease produced by M2 for sufficient time to induce auto-cleavage of NS2/3 protease and produce cleavage products or their fragments, and measuring the presence or absence of uncleaved NS2/3 protease, cleavage products or their fragments ; and (6) screening a potential inhibitor of auto-cleavage activity of an active NS2/3 protease, involving carrying out M3 in the presence of, or absence of, the potential inhibitor, comparing the amount of uncleaved NS2/3 protease, cleavage products or their fragments. The protease is useful for detailed biochemical characterisation of the enzymes and in the development of in vitro assays for screening novel inhibitors of NS2/3 protease which are useful as therapeutic agents against HCV infection (which causes chronic liver disease, cirrhosis and end-stage liver disease). M1 is useful for high level production of protease. The present sequence represents the NS2/3 truncation mutant 855-1206 (numbered relative to the full length NS2/3 protein)

XX SQ Sequence 352 AA;

Query Match 100.0%; Score 1532; DB 5; Length 352;

Best Local Similarity 100.0%; Pred. No. 1.5e-140; Matches 292; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AQLIRACMLYRKAGGHYYQMAFKLALLTGTGVYDHITPLQWAHAGIRDIAVAPV 60

Db 61 AQLIRACMLVRKAGGHYQMAFMKLLAALTGTIVDHTL2LDWAHAGLADLAVAVEPV 120  
 QY 61 IFSDMEVKITWGADTAACGDIISGLPVSARRGREILIGPNDNEGGQRLLAPITAYSQ 120  
 Db 121 IFSDMEVKITWGADTAACGDIISGLPVSARRGREILIGPNDNEGGQRLLAPITAYSQ 180  
 QY 121 QTRGLGCITSLTSRDKNQEGEVQVSTATQSFLATCNGCWTVFHAGSKTLAGPK 180  
 Db 181 QTRGLGCITSLTSRDKNQEGEVQVSTATQSFLATCNGCWTVFHAGSKTLAGPK 240  
 QY 181 GRITQMYTNVDODLVWQAPPGRASMRMPTCTCGSSDLYLVTRHADVIPVRRGDSRSLS 240  
 Db 241 GPITQMYTNVDODLVWQAPPGRASMRMPTCTCGSSDLYLVTRHADVIPVRRGDSRSLS 300  
 QY 241 PRPVSYLKGSSEGPPLCLPSGHAVGIFRAVCTRGVAKAVDFIPVESMETMR 292  
 Db 301 PRPVSYLKGSSEGPPLCLPSGHAVGIFRAVCTRGVAKAVDFIPVESMETMR 352

RESULT 6

ID ABG32185 standard; protein; 380 AA.

XX

AC ABG32185;

XX

DT 05-NOV-2002 (first entry)

XX

DE HCV protease NS2/3 truncation mutant 827-1206.

XX

KW hepatitis C virus infection; chronic liver disease; end-stage liver disease; cirrhosis; virus; hepatotropic agent; mutant; mutein.

XX

OS Synthetic.

OS

PN WO0024375-A2.

XX

PD 20-JUN-2002.

XX

PF 13-DEC-2001; 2001WO-CA001796.

XX

PR 15-DEC-2000; 2000US-0256031P.

XX

PA (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.

XX

PT Thibeault D, Lamatre D, Maurice R, Pilote L, Pause A;

XX

DR WPI; 2002-599511/64.

XX

PT Novel polypeptide for screening inhibitors of non-structural proteases useful as therapeutic agents against hepatitis C virus, comprises full length non-structural protease, or its truncation.

XX

PS Claim 41; Page 60-61; 67PP; English.

XX

CC

PI Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;  
 XX  
 DR  
 WPI; 2002-599511/64.

PT Novel polypeptide for screening inhibitors of non-structural proteases useful as therapeutic agents against hepatitis C virus, comprises full length non-structural protease, or its truncation.  
 PT  
 XX  
 PS Claim 41; Page 59-60; 67pp; English.

XX The invention relates to an isolated polypeptide consisting of a full-length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal residue amino acid 810 to 905, or having a minimal amino acid sequence from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length NS2/3 protease. Also included are (1) a composition (C) comprising an isolated HCV NS2/3 protease selected from full length NS2/3 protease, or its truncation or a mutated sequence, where the protease is in a solution comprising a sufficient concentration of lauryldiethylamine oxide (LDAO) to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3 protease, involving isolating the protease in the presence of a chaotropic agent, refolding the isolated protease by contacting it with a reducing agent, and LDAO in the presence of reduced concentration of the chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3 protease, involving diluting refolded inactive NS2/3 protease in a medium containing an activation detergent to induce auto-cleavage of the NS2/3 protease; (5) measuring (M3) the auto-cleavage activity of NS2/3 protease, involving incubating the active NS2/3 protease produced by M2 for sufficient time to induce auto-cleavage of NS2/3 protease and produce cleavage products or their fragments, and measuring the presence or absence of uncleaved NS2/3 protease, cleavage products or their fragments ; and (6) screening a potential inhibitor of auto-cleavage activity of an active NS2/3 protease, involving carrying out M3 in the presence of, or absence of the potential inhibitor, comparing the amount of uncleaved NS2/3 protease, cleavage products or their fragments. The protease is useful for detailed biochemical characterisation of the enzymes and in the development of in vitro assays for screening novel inhibitors of NS2/3 protease which are useful as therapeutic agents against HCV infection (which causes chronic liver disease, cirrhosis and end-stage liver disease). M1 is useful for high level production of protease. The present sequence represents the NS2/3 truncation mutant 815-1206 (numbered relative to the full length NS2/3 protein)

CC Sequence 393 AA;

XX Query Match 100.0%; Score 1532; DB 5; Length 393;  
 Matches 202; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AQGLIRACMVLRKAAGHVVQMAFMKLAUTGTVYDHLPDQWDWAHAGRLAVAEPV 60  
 DB 102 AQGLIRACMVLRKAAGHVVQMAFMKLAUTGTVYDHLPDQWDWAHAGRLAVAEPV 161  
 OY 61 IFSDEMEVKITITWGAATACGGDINGLPVPSARRREELIGPAPNFEQGWRLLAPIAYSQ 120  
 DB 162 IFSDEMEVKITITWGAATACGGDINGLPVPSARRREELIGPAPNFEQGWRLLAPIAYSQ 221  
 OY 121 QTRGLGCTTSLTGRDKQVEBQVQSTASPLATCNGVQWTFHGAGSKTLGPK 180  
 DB 222 QTRGLGCTTSLTGRDKQVEBQVQSTASPLATCNGVQWTFHGAGSKTLGPK 281  
 OY 181 GPTOMTYTDQDLVQWPGRGSMPCICCGSSDLIVTRHADIPVRERDSRSLLS 240  
 DB 282 GPTOMTYTDQDLVQWPGRGSMPCICCGSSDLIVTRHADIPVRERDSRSLLS 341  
 OY 241 PRPVSYLKGSQGGLPCLPSGHAVGIFRAAVCTRGVAKADFIPVESMETTR 292  
 DB 342 PRPVSYLKGSQGGLPCLPSGHAVGIFRAAVCTRGVAKADFIPVESMETTR 393

XX RESULT 8  
 ABG32181

ID ABG32181 standard; protein; 409 AA.  
 XX  
 AC ABG32181;  
 XX DT 05-NOV-2002 (first entry)  
 DE XX HCV protease NS2/3 (810-1206).  
 DE XX KW HCV; enzyme; protease; NS2/3 (810-1206); hepatitis C virus infection; chronic liver disease; cirrhosis; end-stage liver disease; viricide; hepatotropic; antiinflammatory; lauryldiethylamine oxide; LDAO; chaotropic agent; mutant; mutein.  
 DE XX OS Hepatitis C virus.  
 DE XX Synthentic.

XX FH Key Location/Qualifiers  
 FT Peptide 398. .409  
 FT /note= "Streptavidin tag"  
 XX PN WO200248375-A2.  
 XX PD 20-JUN-2002.  
 XX PR 13-DEC-2001; 2001WO-CA001796.  
 XX PR 15-DEC-2000; 2000US-0256031P.  
 XX PA (BOehringer Ingelheim Canada Ltd.  
 XX PI Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;  
 XX DR WPI; 2002-599511/64.  
 XX DR N-PSDB; ABG9006.  
 XX  
 CC Novel polypeptide for screening inhibitors of non-structural proteases useful as therapeutic agents against hepatitis C virus, comprises full length non-structural protease, or its truncation.  
 XX PS Claim 42; Fig 1B; 67pp; English.

XX The invention relates to an isolated polypeptide consisting of a full-length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal residue amino acid 810 to 905, or having a minimal amino acid sequence from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length NS2/3 protease. Also included are (1) a composition (C) comprising an isolated HCV NS2/3 protease selected from full length NS2/3 protease, or its truncation or a mutated sequence, where the protease is in a solution comprising a sufficient concentration of lauryldiethylamine oxide (LDAO) to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3 protease, involving isolating the protease in the presence of a chaotropic agent, refolding the isolated protease by contacting it with a reducing agent, and LDAO in the presence of reduced concentration of the chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3 protease, involving diluting refolded inactive NS2/3 protease in a medium containing an activation detergent to induce auto-cleavage of the NS2/3 protease; (5) measuring (M3) the auto-cleavage activity of NS2/3 protease, involving incubating the active NS2/3 protease produced by M2 for sufficient time to induce auto-cleavage of NS2/3 protease and produce cleavage products or their fragments, and measuring the presence or absence of uncleaved NS2/3 protease, cleavage products or their fragments ; and (6) screening a potential inhibitor of auto-cleavage activity of an active NS2/3 protease, involving carrying out M3 in the presence of, or absence of the potential inhibitor, comparing the amount of uncleaved NS2/3 protease, cleavage products or their fragments. The protease is useful for detailed biochemical characterisation of the enzymes and in the development of in vitro assays for screening novel inhibitors of NS2/3 protease which are useful as therapeutic agents against HCV infection (which causes chronic liver disease, cirrhosis and end-stage liver disease). M1 is useful for high level production of protease. The present sequence represents the NS2/3 (810-1206) protein, which has a C-

CC terminal streptavidin tag  
 XX Sequence 409 AA;  
 SQ Query Match 100.0%; Score 1532; DB 5; Length 409;  
 Best Local Similarity 100.0%; Pred. No. 1.9e-140; Indels 0; Gaps 0;  
 Matches 292; Conservative 0; Mismatches 0; InDel 0; Gaps 0;  
 QY 1 AQLIRACMLVRKAAGGHYVQMAFMKLAALTGTIVYDHLTPQDWAHAGRLAVAPEV 60  
 Db 106 AQLIRACMLVRKAAGGHYVQMAFMKLAALTGTIVYDHLTPQDWAHAGRLAVAPEV 71  
 QY 61 IFSDEMEVKITWGADTAAGGDSIGLPSARSGRETEILGPADNFEQQWRLLAPITASQ 120  
 Db 166 IFSDEMEVKITWGADTAAGGDSIGLPSARSGRETEILGPADNFEQQWRLLAPITASQ 225  
 QY 121 QTRGLLGCGITSLTGDRKNQVEGEVQVSTATOSFLATCNGVWCFHGAQSKTLAGPK 180  
 Db 286 GRITQMYTNVDQDLVGWQAPPGARSMTPTCGSSDLYLVTRADVIPRRRGSRSLLS 345  
 QY 241 PRPVSYLKGSSEGGPLCPSGHAVGIRAVCTRGVAKAVDFIPVESMETMR 292  
 Db 346 PRPVSYLKGSSEGGPLCPSGHAVGIRAVCTRGVAKAVDFIPVESMETMR 397  
 RESULT 9  
 ABG32191  
 ID ABG32191 standard; protein; 303 AA.  
 AC ABG32191;  
 XX  
 DT 05-NOV-2002 (first entry)  
 XX HCV protease NS2/3 truncation 904-1206/Cys993Ala.  
 DE HCV; enzyme; protease; NS2/3 (904-1206); hepatitis C virus infection; chronic liver disease; cirrosis; end-stage liver disease; virucide; hepatotrophic; antiinflammatory; lauryldiethylamine oxide; LDRO; chaotropic agent; mutant; mutein.  
 XX Hepatitis C virus.  
 OS Hepatitis C virus.  
 XX  
 FH Key Difference 90  
 FT Misc-difference /note= "Wild-type Cys substituted by Ala"  
 XX WO200248375-A2.  
 XX 20-JUN-2002.  
 XX  
 PR 13-DRC-2001; 2001WO-CA001796.  
 XX PR 15-DEC-2000; 2000US-0256031P.  
 XX  
 PA (BOEH ) BOHRINGER INGELHEIM CANADA LTD.  
 PI Thibeault D, Lamerre D, Maurice R, Pilote L, Pause A;  
 XX WPI; 2002-599511/64.  
 PT Novel polypeptide for screening inhibitors of non-structural proteases useful as therapeutic agents against hepatitis C virus, comprises full length non-structural protease, or its truncation.  
 XX Disclosure; Page 65-66; 67pp; English.  
 CC The invention relates to an isolated polypeptide consisting of a full-length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred  
 CC to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal residue amino acid 810 to 906, or having a minimal amino acid sequence from residues 904 to 1205 of hepatitis C virus (HCV) 1b-40 full-length NS2/3 protease. Also included are (1) a composition (C) comprising an isolated HCV NS2/3 protease selected from full length NS2/3 protease, or comprising a sufficient concentration of lauryldiethylamine oxide (LDAO) to prevent auto-clavage of the protease; (2) a NS2/3 inhibitory peptide appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3 protease, involving isolating the protease in the presence of a chaotropic agent, refolding the isolated protease by contacting it with a reducing agent, and LDAO in the presence of reduced concentration of the chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3 protease, involving diluting refolded, inactive NS2/3 protease in a medium containing an activation detergent to induce auto-cleavage of the NS2/3 protease; (5) measuring (M3) the auto-cleavage activity of NS2/3 protease, involving incubating the active NS2/3 protease produced by M2 for sufficient time to induce auto-cleavage of NS2/3 protease and produce cleavage products or their fragments, and measuring the presence or absence of uncleaved NS2/3 protease, cleavage products or their fragments; and (6) screening a potential inhibitor of auto-cleavage activity of an active NS2/3 protease, involving carrying out M3 in the presence of, or absence of the potential inhibitor, comparing the amount of uncleaved NS2/3 protease, cleavage products or their fragments. The protease is useful for detailed biochemical characterisation of the enzymes and in the development of in vitro assays for screening novel inhibitors of NS2/3 protease which are useful as therapeutic agents against HCV infection (which causes chronic liver disease, cirrhosis and end-stage liver disease). M1 is useful for high-level production of protease. The present sequence represents the NS2/3 truncation 904-1206 mutant (Cys993Ala (numbered relative to the full length NS2/3 protein) a mutant devoid of autocatalytic activity  
 XX Sequence 303 AA;  
 SQ Query Match 99.4%; Score 1523; DB 5; Length 303;  
 Best Local Similarity 99.7%; Pred. No. 9.3e-140; Indels 1; Gaps 0; Gaps 0;  
 Matches 291; Conservative 0; Mismatches 1; InDel 0; Gaps 0;  
 QY 1 AQLIRACMLVRKAAGGHYVQMAFMKLAALTGTIVYDHLTPQDWAHAGRLAVAPEV 60  
 Db 12 AQLIRACMLVRKAAGGHYVQMAFMKLAALTGTIVYDHLTPQDWAHAGRLAVAPEV 71  
 QY 61 IFSDEMEVKITWGADTAAGGDSIGLPSARSGRETEILGPADNFEQQWRLLAPITASQ 120  
 Db 72 IFSDEMEVKITWGADTAAGGDSIGLPSARSGRETEILGPADNFEQQWRLLAPITASQ 131  
 QY 121 QTRGLLGCGITSLTGDRKNQVEGEVQVSTATOSFLATCNGVWCFHGAQSKTLAGPK 180  
 Db 132 QTRGLLGCGITSLTGDRKNQVEGEVQVSTATOSFLATCNGVWCFHGAQSKTLAGPK 191  
 QY 181 GPTQMYTNVDQDLVGWQAPPGARSMTPTCGSSDLYLVTRADVIPRRRGSRSLLS 240  
 Db 192 GRITQMYTNVDQDLVGWQAPPGARSMTPTCGSSDLYLVTRADVIPRRRGSRSLLS 251  
 QY 241 PRPVSYLKGSSEGGPLCPSGHAVGIRAVCTRGVAKAVDFIPVESMETMR 292  
 Db 252 PRPVSYLKGSSEGGPLCPSGHAVGIRAVCTRGVAKAVDFIPVESMETMR 303  
 RESULT 10  
 ABG32199  
 ID ABG32199 standard; protein; 303 AA.  
 XX  
 AC ABG32199;  
 XX  
 DT 05-NOV-2002 (first entry)  
 XX HCV protease NS2/3 truncation 904-1206/His952Ala.  
 XX  
 PR HCV; enzyme; protease; NS2/3 (904-1206); hepatitis C virus infection; chronic liver disease; cirrosis; end-stage liver disease; virucide; hepatotrophic; antiinflammatory; lauryldiethylamine oxide; LDRO;

|           |   |
|-----------|---|
| KW        | chaotropic agent; mutant; mutein.   |
| XX        |   |
| OS        | Hepatitis C virus.  |
| XX        | Synthetic.  |
| Key       | Location/Qualifiers   |
| FT        | Misc-difference 49 /note= "Wild-type His substituted by Ala"  |
| XX        | WO200248375-A2.   |
| PN        | XX  |
| DR        | 20-JUN-2002.  |
| PD        |   |
| XX        | 13-DEC-2001; 2001WO-CA001796.   |
| PP        | PR  |
| XX        | 15-DEC-2000; 2000US-0256031P.   |
| PA        | (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  |
| XX        |   |
| PI        | Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;   |
| XX        |   |
| DR        | WPI; 2002-599511/64.  |
| XX        |   |
| PT        | Novel polypeptide for screening inhibitors of non-structural proteases useful as therapeutic agents against hepatitis C virus, comprises full length non-structural protease, or its truncation.  |
| PS        | Example 7; Fig 8; 67pp; English.  |
| XX        | The invention relates to an isolated polypeptide consisting of a full-length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal residue amino acid 810 to 906, or having a minimal amino acid sequence from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length NS2/3 protease. Also included are (1) a composition (C) comprising an isolated HCV NS2/3 protease selected from full length NS2/3 protease, or its truncation or a mutated sequence where the protease is in a solution comprising a sufficient concentration of lauryldiethylamine oxide (LDAO) to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide appearing as ABG32198; (3) producing (M1) a refolded, inactive HCV NS2/3 protease, involving isolating the protease in the presence of a chaotropic agent, refolding the isolated protease by contacting it with a reducing agent, and LDAO in the presence of reduced concentration of the chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3 protease, involving diluting refolded inactive NS2/3 protease in a medium containing an activation detergent to induce auto-cleavage of the NS2/3 protease; (5) measuring (M3) the auto-cleavage activity of NS2/3 protease, involving incubating the active NS2/3 protease produced by M2 for sufficient time to induce auto-cleavage of NS2/3 protease and produce cleavage products or their fragments, and measuring the presence or absence of uncleaved NS2/3 protease, cleavage products or their fragments ; and (6) screening a potential inhibitor of auto-cleavage activity of an active NS2/3 protease, involving carrying out M3 in the presence of, or absence of the potential inhibitor, comparing the amount of uncleaved NS2/3 protease, cleavage products or their fragments. The protease is useful for detailed biochemical characterisation of the enzymes and in the development of in vitro assays for screening novel inhibitors of NS2/3 protease which are useful as therapeutic agents against HCV infection (which causes chronic liver disease, cirrhosis and end-stage liver disease). It is useful for high level production of protease. The present sequence represents the NS2/3 truncation 904-1206 mutant devoid of autocatalytic activity |
| CC        | Sequence 303 AA;  |
| CC        | Query Match 99.3%; Score 1522; DB 5; Length 303;  |
| CC        | Best Local Similarity 99.7%; Pred. No. 1.2e-139; Matches 291; Conservative 0; Mismatches 1; Indels 0; Gaps 0;   |
| Ov        | 1 AGQILRACMLVRLKAGGHIVQAFMKLAUTGTYVPHITLQDWAHQGLDVLAVEPV 60   |
| Db        | 12 AQGLIRACMLVRLKAGGHIVQAFMKLAUTGTYVPHITLQDWAHQGLDVLAVEPV 71  |
| Qy        | 61 IFSDMEVKITWGADTAACGDIISGLPVSVARRGRETLIGPADNFEGQQRLLAPITAYSQ 120  |
| Db        | 72 IFSDMEVKITWGADTAACGDIISGLPVSVARRGRETLIGPADNFEGQQRLLAPITAYQ 131   |
| Qy        | 121 QTRGLIQCIIITSITGRDKNOVEGEYQVSVSTATQSFLATCTNGVCTVFHAGSKTLAGPK 180  |
| Db        | 132 QTRGLIQCIIITSITGRDKNOVEGEYQVSVSTATQSFLATCTNGVCTVFHAGSKTLAGPK 191  |
| Qy        | 181 GPTQMTNTQDQGWAQPGARGMTCTCGSSDLVYTHADVIPVERRGSGRSLS 240  |
| Db        | 192 GPTQMTNTQDQGWAQPGARGMTCTCGSSDLVYTHADVIPVERRGSGRSLS 251  |
| Qy        | 241 PRPVSIKGSSGGPUCPSGHAVGPRAAVCTRGAKADPIPEMESMTMR 292  |
| Db        | 252 PRPVSIKGSSGGPUCPSGHAVGPRAAVCTRGAKADPIPEMESMTMR 303  |
| RESULT 11 |   |
| ID        | ABG32190  |
| XX        | ABG32190 standard; protein; 301 AA.   |
| XX        |   |
| AC        | ABG32190;   |
| XX        |   |
| DT        | 05-NOV-2002 (first entry)   |
| XX        |   |
| DE        | HCV protease NS2/3 truncation 904-1205/deltaLeu1026-Ala1027.  |
| XX        |   |
| KW        | HCV; enzyme; protease; NS2/3 (904-1205); hepatitis C virus infection; chronic liver disease; cirrhosis; end-stage liver disease; virucide; hepatocropic; antiinflammatory; lauryldiethylamine oxide; LDAO; chaotropic agent; mutant; mutein.  |
| XX        |   |
| OS        | Hepatitis C virus.  |
| OS        | Synthetic.  |
| XX        |   |
| PH        |   |
| XX        |   |
| PH        | Misc-difference 122. 123 /note= "Wild-type Leu-Leu-Ala-Pro substituted by Leu-Pro"  |
| XX        |   |
| XX        | WO200248375-A2.   |
| XX        |   |
| PD        | 20-JUN-2002.  |
| XX        |   |
| PP        | 13-DEC-2001; 2001WO-CA001796.   |
| XX        |   |
| PR        | 15-DEC-2000; 2000US-0256031P.   |
| XX        |   |
| PA        | (BOEH ) BOEHRINGER INGELHEIM CANADA LTD.  |
| XX        |   |
| PI        | Thibeault D, Lamarre D, Maurice R, Pilote L, Pause A;   |
| XX        |   |
| DR        | WPI; 2002-599511/64.  |
| XX        |   |
| PT        | Novel polypeptide for screening inhibitors of non-structural proteases useful as therapeutic agents against hepatitis C virus, comprises full length non-structural protease, or its truncation.  |
| PS        | Example 7; page 64-65; 67pp; English.   |
| XX        | The invention relates to an isolated polypeptide consisting of a full-length HCV (hepatitis C virus) non-structural (NS)2/3 protease (referred to also as NS2/3 (810-1206)), or its truncation, having as its N-terminal residue amino acid 810 to 906, or having a minimal amino acid sequence from residues 904 to 1206 of hepatitis C virus (HCV) 1b-40 full-length NS2/3 protease. Also included are (1) a composition (C) comprising an isolated HCV NS2/3 protease selected from full length NS2/3 protease, or its truncation or a mutated sequence, where the protease is in a solution comprising a sufficient concentration of lauryldiethylamine oxide (LDAO) to prevent auto-cleavage of the protease; (2) a NS2/3 inhibitory peptide appearing as ABG3198; (3) producing (M1) a refolded, inactive HCV NS2/3 protease, involving isolating the protease in the presence of a chaotropic agent, refolding the isolated protease by contacting it with a   |

CC reducing agent, and IDAO in the presence of reduced concentration of the  
 CC chaotropic agent or a polar additive; (4) producing (M2) an active NS2/3  
 CC protease, involving diluting refolded inactive NS2/3 protease in a medium  
 containing an activation detergent to induce auto-cleavage activity of the NS2/3  
 protease; (5) measuring (M3) the auto-cleavage activity of NS2/3  
 protease, involving incubating the active NS2/3 protease produced by M2  
 for sufficient time to induce auto-cleavage of NS2/3 protease and produce  
 cleavage products or their fragments, and measuring the presence or  
 absence of uncleaved NS2/3 protease, cleavage products or their fragments  
 ; and (6) screening a potential inhibitor of auto-cleavage activity of an  
 active NS2/3 protease, involving carrying out M3 in the presence of, or  
 absence of the potential inhibitor comparing the amount of uncleaved  
 NS2/3 protease, cleavage products or their fragments. The protease is  
 used for detailed biochemical characterisation of the enzymes and in  
 the development of in vitro assays for screening novel inhibitors of  
 NS2/3 proteases which are useful as therapeutic agents against HCV  
 infection (which causes chronic liver disease, cirrhosis and end-stage  
 liver disease. M1 is useful for high level production of protease. The  
 present sequence represents the NS2/3 truncation 94-106 mutant  
 deltaLeu1026-Ala1027 (numbered relative to the full length NS2/3 protein)  
 a mutant devoid of autocatalytic activity

XX Sequence 301 AA;

Query Match 98.8%; Score 1513; DB 5; Length 301;  
 Best Local Similarity 99.3%; Pred. No. 8.7e-139; Matches 290; Conservatve 0; Mismatches 0; Indels 2; Gaps 1;

QY 1 A Q G L I R A C M L V R K A A G H Y V Q M A F M K L A L T G Y V D Y H L T P Q D W A H G L R L A V E P V 60  
 Db 12 A Q G L I R A C M L V R K A A G H Y V Q M A F M K L A L T G Y V D Y H L T P Q D W A H G L R L A V E P V 71  
 QY 61 I F S D M E V K I T W G A D T A C G D I S G L P S A R G R E G E B I L G P A D N F E G Q C W R L L A P I T A Y S Q 120  
 Db 72 I F S D M E V K I T W G A D T A C G D I S G L P S A R G R E G E B I L G P A D N F E G Q C W R L L A P I T A Y S Q 129  
 QY 121 O T R G L I G C I T S L T G R D K N Q V E G E V Q V V S T A T O S F L A T C V N G Y C W T V F H G A G S K T I L A G P K 180  
 Db 130 Q T R G L I G C I T S L T G R D K N Q V E G E V Q V V S T A T O S F L A T C V N G Y C W T V F H G A G S K T I L A G P K 189  
 QY 181 G P I T O M Y T N V D Q D L V G W Q A P P G A R S M T P C T C G S S D L Y V T R H A D V I P V R R G D R G S L S 240  
 Db 190 G P I T O M Y T N V D Q D L V G W Q A P P G A R S M T P C T C G S S D L Y V T R H A D V I P V R R G D R G S L S 249  
 QY 241 P R P V S Y L K G S S G G P L C I C P S G H A V G I F R A V C T R G V A K A V D F I P V E S M E T T M R 292  
 Db 250 P R P V S Y L K G S S G G P L C I C P S G H A V G I F R A V C T R G V A K A V D F I P V E S M E T T M R 301

XX Sequence 768 AA;

Query Match 96.7%; Score 1481; DB 2; Length 768;  
 Best Local Similarity 95.2%; Pred. No. 4.3e-135; Matches 278; Conservatve 8; Mismatches 6; Indels 0; Gaps 0;

QY 1 A Q G L I R A C M L V R K A A G H Y V Q M A F M K L A L T G Y V D Y H L T P Q D W A H G L R L A V E P V 60  
 Db 176 A Q G L I R A C M L V R K A A G H Y V Q M A F M K L A L T G Y V D Y H L T P Q D W A H G L R L A V E P V 235  
 QY 61 I F S D M E V K I T W G A D T A C G D I S G L P S A R G R E G E B I L G P A D N F E G Q C W R L L A P I T A Y S Q 120  
 Db 236 V F S A M E V K I T W G A D T A C G D I S G L P S A R G R E G E B I L G P A D N F E G Q C W R L L A P I T A Y S Q 295  
 QY 121 O T R G L I G C I T S L T G R D K N Q V E G E V Q V V S T A T O S F L A T C V N G Y C W T V F H G A G S K T I L A G P K 180  
 Db 296 Q T R G L I G C I T S L T G R D K N Q V E G E V Q V V S T A T O S F L A T C V N G Y C W T V F H G A G S K T I L A G P K 355  
 QY 181 G P I T O M Y T N V D Q D L V G W Q A P P G A R S M T P C T C G S S D L Y V T R H A D V I P V R R G D R G S L S 240  
 Db 356 G P I T O M Y T N V D Q D L V G W Q A P P G A R S L T P C T C G S S D L Y V T R H A D V I P V R R G D R G S L S 415  
 QY 241 P R P V S Y L K G S S G G P L C I C P S G H A V G I F R A V C T R G V A K A V D F I P V E S M E T T M R 292  
 Db 416 P R P V S Y L K G S S G G P L C I C P S G H A V G I F R A V C T R G V A K A V D F I P V E S M E T T M R 467

XX Sequence 13 AA;

Query Match 96.7%; Score 1481; DB 2; Length 13;  
 Best Local Similarity 95.2%; Pred. No. 4.3e-135; Matches 13; Conservatve 8; Mismatches 6; Indels 0; Gaps 0;

RESULT 12  
 AAR40223  
 ID AAR40223 standard; protein; 768 AA.  
 AC AAR40223;  
 XX  
 DT 21-FEB-1994 (first entry)  
 DE Recombinant hepatitis C virus genomic protein.  
 XX  
 KW Hepatitis C virus.  
 FH Location/Qualifiers  
 Key  
 PT Misc-difference 10 /note= "UUA encodes lle."  
 FT Misc-difference 81..82 /note= "Nucleotide sequence encodes another Gly"  
 FT Duplication 528..768 /note= "Duplication of 241 amino acids at start of protein sequence."  
 FT Misc-difference 537.. /note= "UUA encodes lle."  
 XX

RESULT 13  
 AAE20477  
 ID AAE20477 standard; protein; 3010 AA.  
 AC AAE20477;  
 XX  
 DT 01-JUL-2002 (first entry)  
 DE HCV-S1 full-length polyprotein.  
 XX  
 KW Nucleic acid construct; expression cassette; non-coding region; NCR;  
 KW untranslated region; UTR; anti-viral drug; drug resistance; HCV-S1;  
 KW Hepatitis C virus.  
 OS Hepatitis C virus.  
 PN W0200208447-A2.  
 XX  
 PD 31-JAN-2002.  
 XX  
 PF 20-JUL-2001; 2001WO-II000669.



RESULT 15

ABG30591  
ID ABG30591 standard; protein; 2201 AA.  
XX  
AC ABG30591;  
XX  
DT 21-OCT-2002 (first entry)  
XX  
DE Hepatitis C virus NS2/3, NS3/4, NS3 and NSSB mutant #3.  
XX  
KW Self-replicating; hepatitis C virus; HCV; HCV replication inhibitor;  
cell culture replication; NS2/3; NS3/4; NS3; NSSB; mutant; muclein.  
XX  
OS Hepatitis C virus.  
Synthetic.

XX  
FH Location/Qualifiers  
Key  
FT Misc-difference 751 /note= "wild type Ser substituted by Gly"  
FT Misc-difference 882 /label= Arg, Lys  
XX  
PN WO200252015-A2.  
XX  
PD 04-JUL-2002.  
XX  
PP 20-DEC-2001; 2001WQ-CA001843.  
PR 22-DEC-2000; 2000058-0257857P.  
XX  
PA (BOEHRINGER INGELHEIM CANADA LTD.  
XX  
PI Kukolj G, Pause A;  
XX  
DR WPI, 2002-575382/61.  
XX  
PT New self-replicating RNA molecules from Hepatitis C virus (HCV), which  
possess enhanced transduction or replication efficiency, useful for  
evaluating potential inhibitors of HCV replication.  
XX  
PS Claim 3; Page: 140pp; English.

The invention describes a self-replicating hepatitis C virus (HCV) poly nucleotide molecule comprising a 5'-non translated region (NTR), where guanine at position 1 is substituted for adenine, a HCV polyprotein region coding for a HCV polyprotein, and a 3'-NTR region. The self-replicating Hepatitis C virus (HCV) RNA molecule is useful for evaluating potential inhibitors of HCV replication. The HCV RNA molecule is also useful for efficiently establishing cell culture replication. The self-replicating poly nucleotide molecule contains a 5'-NTR, where G at position 1 is substituted for A, and therefore provides an alternative to existing systems comprising a self-replicating HCV RNA molecule that, in conjunction with mutations in the HCV non-structural region, such as the G1(2042)C/R mutations, transduces and/or replicates with greater efficiency. This amino acid sequence represents a mutant of the hepatitis C virus replicon APGK12 and contains the viral protease NS2/3, protease complex NS3/4, helicase NS3 and RNA-dependent RNA polymerase NSSB. Note: This sequence does not appear in the specification but has been created from the wild type sequence shown in ABG30580 using information given in the claims of the invention.

XX  
SQ Sequence 2201 AA;

Query Match 96.1%; Score 1473; DB 5; Length 2201;  
Best Local Similarity 94.9%; Pred. No. 1.1e-133;  
Matches 277; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY 1 AQGLURACMLVKRAAGGHYVOMAFMKLALLTGTYYDHTPLQDWAGLERDLAVAEPV 60  
105 AHGLURACMLVKRAAGGHYVOMAFMKLALLTGTYYDHTPLQDWAGLERDLAVAEPV 165  
Db 61 IFSDNEVKITWGADTAACGDIISGILPVSARRGRETILLGPADNFGQGWRILLAPTTAYSQ 120

Db 166 VFSDMETKVITWGADFAAGGIIILGPVPSARGREHLGPASLGEGQWRIAPITAYQ 225  
121 QTRGLIGCITISLTGRDKNQTEGEVQVSTATOSFLATCVNCVQCVWTVFHGASKTLAGPK 180  
226 QTRGULGCIISLTGRDNQTEGEVQVSTATOSFLATCVNCVQCVWTVFHGASKTLAGPK 285  
QY 181 QPITOYNTNVDQDLYWQAPPGRSMPTCGSSDPLVTHADVPPRRGDSRSLLS 240  
286 GPITOYNTNVDQDLYWQAPPGRSLPTCTGSSDPLVTRHADVPPRRGDSRSLLS 345  
Db 241 PRPPSVYKGSSGGPLCPSPGHAVGIRAAVUTRGVAKAVDFPVESETMR 292  
Db 346 PRPPSVYKGSSGGPLCPSPGHAVGIFRAAVUTRGVAKAVDFPVESETMR 397

Search completed: May 6, 2004, 09:30:49  
Job time : 40.7461 secs

